

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:45 ; Search time 10.3976 Seconds  
(without alignments)  
1331.870 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVTVTIQNGKMSSTIVSE.....ATVLDKNISSKSTNNPNK 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	748	100.0	2140	F95074	serine proteinase,
2	745	99.6	2144	A97942	metalloproteinase,
3	110	14.7	558	T18467	hypothetical prote
4	108.5	14.5	1038	JC5497	claustrin - chicke
5	104	13.9	665	B71609	hypothetical prote
6	103.5	13.8	325	T18283	hypothetical prote
7	100	13.4	312	G81339	probable membrane
8	99	13.2	211	T25911	hypothetical prote
9	97.5	13.0	1345	S46817	hypothetical prote
10	97	13.0	988	T14188	hypothetical prote
11	95.5	12.8	644	T47835	cylicin ii - human
12	95	12.7	348	I37271	hypothetical prote
13	95	12.7	535	T37189	hypothetical prote
14	95	12.7	622	A90570	lipoprotein (impor
15	95	12.7	1397	T10466	DNA topoisomerase
16	94.5	12.6	3724	T18427	hypothetical prote
17	94	12.6	210	T28771	hypothetical prote
18	92.5	12.4	219	B72291	hypothetical prote
19	92.5	12.4	540	D86432	hypothetical prote
20	91	12.2	253	T32879	hypothetical prote
21	91	12.2	385	T20410	hypothetical prote
22	90.5	12.1	614	A84152	hypothetical prote
23	90.5	12.1	670	T28391	ORF MSV230 hypothe
24	90.5	12.1	1332	S41552	probable transcrip
25	90.5	12.1	2401	T28676	rhoetry protein -
26	89.5	12.0	508	B81594	hypothetical prote
27	89.5	12.0	508	C72074	hypothetical prote
28	89.5	12.0	508	E86549	hypothetical prote
29	89.5	12.0	700	S67610	probable membrane

30	89.5	12.0	1016	2	T19006	ankyrin related pr
31	89	11.9	528	2	E96795	unknown protein F2
32	89	11.9	645	2	E89883	conserved hypothet
33	89	11.9	1888	2	T39009	hypothetical prote
34	88.5	11.8	301	2	T33068	hypothetical prote
35	88.5	11.8	371	2	A71683	hypothetical prote
36	88.5	11.8	762	2	G88436	protein T04A8.13 l
37	88.5	11.8	791	2	T24435	hypothetical prote
38	88	11.8	276	2	T23451	hypothetical prote
39	88	11.8	1280	2	G96796	Iga-specific metal
40	87.5	11.7	1702	2	A41859	hypothetical prote
41	87.5	11.7	2523	2	T18477	hypothetical prote
42	87	11.6	635	2	F71621	microtubule-associ
43	87	11.6	2464	1	QRMSP1	probable chloroqui
44	87	11.6	2708	2	T09079	hypothetical prote
45	86.5	11.6	792	2	T49989	hypothetical prote

ALIGNMENTS

RESULT 1

F95074

serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: F95074

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,

nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: F95074

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2140 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK74791.1; PID:G14972117; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0641

Query Match 100.0%; Score 748; DB 2; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 6.3e-49;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQPDGWEISGFEGKDGAGYVNLK 60  
|||||  
Db 1963 HRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQPDGWEISGFEGKDGAGYVNLK 2022

Qy 61 DTFIKPVFKTEKEEENKPTDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 120  
|||||  
Db 2023 DTFIKPVFKTEKEEENKPTDVSKKONPVNHSQLNESHKEDLQREHSQKSDST 2082

Qy 121 KDVTATVLDKNISSKSTNNPNK 144  
|||||  
Db 2083 KDVTATVLDKNISSKSTNNPNK 2106

RESULT 2

A97942

metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001

C:Accession: A97942

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bargett, S.; DeHoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: A97942  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2144 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99365.1; PID:g15458138; GSPDB:GN00174  
C:Genetics:  
A:Gene: prta  
C:Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 745; DB 2; Length 2144;  
Best Local Similarity 99.3%; Pred. No. 1.1e-48;  
Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRVTVTQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVNLISK 60  
DB 1967 HRVTVTQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVNLISK 2026

QY 61 DTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSOKSDST 120  
DB 2027 DTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSOKSDST 2086

QY 121 KDVATVLDKNNISSKSTTNNPNK 144  
DB 2087 KDVATVLDKNNISSKSTTNNPNK 2110

RESULT 3  
T18467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2002  
C:Accession: T18467  
R:Lawson, D.; Bowman, S.; Barrell, B.  
A:Reference number: Z18937  
A:Accession: T18467  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-558 <LAW>  
A:Cross-references: EMBL:AL008970; NID:e1407852; PIDN:CAA15610.1  
C:Genetics:  
A:Map position: 3  
A:Introns: 84/1; 160/1  
A>Note: C0465c

Query Match 14.7%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 0.5;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 17 IVSEEDFILPVY-----KGELEKGYQFDGWEISGFEKK-----DAGYVNLKDTFIKPV 67  
DB 60 ILGFEDDILYCYCISQLKQSKK--KADGEEDKYLNAKKLKNLTGFIGNKKSDIFIEL 117

QY 68 FKKI--EEKKEE-----ENKPTFDVSK-KKDNQVNHSQLNE-----SHRK 105  
DB 118 LELLNEEKKEEHIADTLNENK-TNDIKVKYKNNENINENYNNENKDNKDEHVSQHN 176

QY 106 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISSKSTTN 140  
DB 177 EHNINNVNLKKEKYTDIQDKKKHRSLSQKSDSYKKEPFNKRKTSIER-SLSNKRIDE 235

QY 141 NPNK 144  
DB 236 KTNK 239

RESULT 4  
JC5497  
claustrian - chicken  
N:Alternate names: keratan sulfate proteoglycan  
C:Species: Gallus gallus (chicken)  
C:Date: 07-Jul-1997 #sequence\_revision 12-Sep-1997 #text\_change 21-Jul-2000  
C:Accession: JC5497; PC4334; S37561

R:Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A:Title: Claustrian, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A:Reference number: JC5497; MUID:94157526; PMID:7906711  
A:Accession: JC5497  
A:Molecule type: mRNA  
A:Residues: 1-1038 <BUR1>  
A:Cross-references: EMBL:X67778; NID:g406318; PIDN:CAA47988.1; PID:g406319  
A:Accession: PC4334  
A:Molecule type: protein  
A:Residues: 79-83/299-412/485-502 <BUR2>  
A:Experimental source: brain  
C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nerve  
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
F:267-270/Region: cell attachment (R-G-D) motif  
F:112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 14.5%; Score 108.5; DB 2; Length 1038;  
Best Local Similarity 28.6%; Pred. No. 1.3;  
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

QY 11 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVNLKDTFIKPVFK- 69  
DB 595 KPETKTIVAERDV-----TTKEQLGKSETSEKQASEKQDVKPKVTKKSVKVKKA 646

QY 70 KIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQRE-----EHSQKSDSTKV 123  
DB 647 KPKEKKDEKPKKPKKPKKPKKPLI---KBEKPKKEDIKKEVKKEKKEKKEKKEV 702

RESULT 5  
B71609  
hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: B71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: B71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-665 <GAR>  
A:Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AACT1925.1; PID:g384524  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0680w

Query Match 13.9%; Score 104; DB 2; Length 665;  
Best Local Similarity 25.9%; Pred. No. 1.8;  
Matches 42; Conservative 29; Mismatches 35; Indels 56; Gaps 10;

QY 1 HRVT-VTTQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEI---SGFEKKDAGYVNL 57  
DB 138 HRQNELNLQSGK-----NEQDI-----NKNKRGKQ-----DISNSNAENKCD----- 174

QY 58 LSKDTFIKPVFKKIEKKE-----EENKPTFD-----VSKKDNQVNHSQLNESH 103  
DB 175 -----VRGKVELEKKEKKEKISDDHKVEENKSDDHKVEENKSDDHKVEENKSDDH 228

QY 104 RKEDLQR-EHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 144  
DB 229 KIEEVKKVEEHEDEE-----DKKEKSENKKNKDNK 261

RESULT 6  
T18283  
hypothetical protein G5 - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999





```

C;date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 21-Jul-2000
C;Accession: I37271; S22774
R;Hess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.
Exp. Cell Res. 218, 174-182, 1995
A;Title: The protein complexity of the cytoskeleton of bovine and human sperm h
A;Reference number: I37271; MUID:95255491; PMID:71737358
A;Accession: I37271
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-348 <HES>
A;Cross-references: EMBL:Z46788; NID:g758586; PIDN:CAA86752.1; PID:g758587

Query Match          12.7%  Score 95;  DB 2;  Length 348;
Best Local Similarity 30.9%;  Pred. No. 4.;
Matches 38;  Conservative 20;  Mismatches 39;  Indels 26;  Gaps 7;

Qy      29  KGELENGYQDGWEISGFEFGKDGAGYVNLSDTKFIKPVFKKIEEKENKPTF---DV 85
          :  ||||  :  ||||  :  ||||  :  ||||  :  ||||  :  ||||  :
Db      205  ESSEKGN-----GTEKSKKGGKUS-----KKGKSAIELOAVKADKKDCKDANKGDE 256

Qy      86  SK---KKDNPOVNHSQLN-----ESHKEDLQREHSQKSDSTKD---VTATVLDKNNI 133
          :  ||||  :  ||||  :  ||||  :  ||||  :  ||||  :  ||||  :
Db      257  SKDAKDAKEIKKGKDKKPKSPSTSDSDSKDDVKE---SKKDATKDAKVAKDKTEKSA 313

Qy      134  SSK 136
          ||
Db      314  DSK 316

```

RESULT 13

T37189  
hypothetical protein C02H7.1 - *Caenorhabditis elegans*  
C/Species: *Caenorhabditis elegans*  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000  
C/Accession: T37189  
R/Leimbac, D.; Minx, M.  
submitted to the EMBL Data Library, February 1996  
A/Description: The sequence of *C. elegans* cosmid C02H7.  
A/Reference number: Z20523  
A/Accession: T37189  
A/Status: preliminary; translated from GB/EMBL/DBDJ  
A/Molecule type: DNA  
A/Residues: 1-535 <LEI>  
A/Cross-references: EMBL:U49945; PIDN:AAC47924.1; GSPDB:GN00029; CESP:C02H7.1  
A/Experimental source: strain Bristol N2; clone C02H7  
C/Genetics:  
A/Gene: CESP:C02H7.1  
A/Map position: X  
A/Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

[illegible]

A90570  
lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
C:Species: Mycoplasma pulmonis  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
C:Accession: A90570

A;Accession: A90570  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-622 <KUR>  
A;Cross-references: GB:AL45566; PID:gl4089879; PIDN:CAC13638.1; GSPDB:GNO0153  
A;Experimental source: strain UAB CT1P  
C;Genetics:  
A;Gene: MYPU\_4650  
A;Genetic code: SGC3

Query Match	12.7%	Score 95	DB 2	Length 622
Best Local Similarity	25.4%	Pred. No. 7.9		
Matches	34	Conservative 26	Mismatches 50	Indels 24
Gaps	5			
QY	18	VSEDFILPVYKGLKGYQFDGWEISGFEGKDGAGVYNLSKDTFFKVPFKTIEEKKE	77	
DB	85	LSEDNIIFSILNTEINNAKQLDEF-----VSKDEKPKIQEINFSQTE	128	
QY	78	ENKPTFDVSKKD-----NPQVHSQLNES--HRKEDLQREHSQKSDSTKDVTATVLDKN	131	
DB	129	Q-KITDNISSKEDEKNPKPDNENNNSSDPQKNDLQKNNSDKLNDVQDEKANKENS	187	
QY	132	-NISSSKSTNNPNK	144	
DB	188	SNDSKEKNDENTNK	201	

RESULT 15  
T10466  
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.13) II - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000  
C:Accession: T10466  
R:Cheeseman, S.J.  
submitted to the EMBL Data Library, September 1995  
A:Reference number: Z17031  
A:Accession: T10466  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1397 <CHE>  
A:Cross-references: EMBL:X79345; NID:G994807  
C:Genetics:  
A:Gene: TopoII  
A:Map position: 14  
C:Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hydrolyzing)  
C:Keywords: ATP; DNA binding; isomerase; nucleus

Query Match	12.7%;	Score 95;	DB 2;	Length 1397;
Best Local Similarity	23.8%;	Pred. No. 20;		
Matches 36;	Conservative 36;	Mismatches 53;	Indels 26;	Gaps 6;
Qy	9	NGKEMSTIYSEEDFIL--PYVKGELSKGYQFGDGEISGREGKDGAVYNLSKDTFKP	66	
Db	1132	NEHIIAGITVKDYLLSMPIFSUTLEK---VEOLLQLKERELEILRNITVETWMLK	1188	
Qy	67	VFKKIEE-----KKEEENKPTFDVSKKDNQPVNHSQNLNESHRKEDLQREHSQ	115	
Db	1189	DIEKVEEAIEFQRNVELSNREESNK--FKVARKQ-----GPSSMKKKKKKLSDEESE	1241	
Qy	116	---KSSTKQVATATVLDKNNISSKSTNNPN	143	
Db	1242	GGTSDSSSEFLVNTLNKKQNTKKTTSSNN	1272	

RESULT 16  
T18427  
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text\_change 09-Jun-2000

C;Accession: T18427  
R;Lawson, D.; Bowman, S.; Barrel, B.  
submitted to the EMBL Data Library, August 1997  
A;Reference number: Z18935  
A;Accession: T18427  
A;Status: preliminary; translated from GB/EMBL/DBDJ  
A;Molecule type: DNA  
A;Residues: 1-3724 <LAW>  
A;Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1  
C;Genetics:  
A;Introns: 307/1; 1545/2  
A;Note: C0335C

Query Match	12.6%;	Score 94.5;	DB 2;	Length 3724;
Best Local Similarity	22.1%;	Pred. No. 64;		
Matches 32;	Conservative 30;	Mismatches 50;	Indels 33;	Gaps 6;
Qy	1	HRVTVTIQNGKMSSTIVSEEDFILPYVKGELBKGQFDCGWEISGPEGKKGAGVYNLSK	60	
Db	1046	HKTEQIDIQHSIQTNICDNN-IQGINEENSKGVRISGDM---ENKND-----	1092	
Qy	61	DTPIKPVFKIEBKSEENKPTPDVSKKDNQVNH-SQLNESHKVEDLQREHSQKSDS	119	
Db	1093	-----MENKNDMEKGN--DMEKNDIEKKNDMEKGNDEKKNDEKKNDEKKNDEKKN	1140	
Qy	120	TKDVTATVLDKNNISSKSTTNPNK	144	
Db	1141	ME-----NENNMENKSDIENENK	1158	

RESULT 17

T28771  
hypothetical protein E03H12.5 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T28771  
R;Nelson, J.; Wohldmann, P.; Sansone, J.  
submitted to the EMBL Data Library, June 1997  
A;Description: The sequence of *C. elegans* cosmid E03H12.  
A;Reference number: Z20520  
A;Accession: T28771  
A;Status: preliminary; translated from GB/EMBL/DD5J  
A;Molecule type: DNA  
A;Residues: 1-210 <NBL>  
A;Cross-references: EMBL:AF002099; PIDN:AA047980.1; GSPDB:GNC00022; CESP:E03H12.  
A;Experimental source: strain Bristol N2; clone E03H12  
C;Genetics:  
A;Gene: CESP:E03H12.5  
A;Map position: 4  
A;Introns: 30/2; 201/3

	Query Match	12.6%;	Score 94;	DB 2;	Length 210;	
	Best Local Similarity	28.6%;	Pred. No. 2.8;			
	Matches	28;	Conservative 18;	Mismatches 52;	Indels 0;	Gaps 0;
Qy	47	EKKDAGYVNLISKDTFIPVPVKIIEKKEEENKPTFDYSKKDNQVNHSQLNESHRKE	106			
		:   :	:   :	:   :	:   :	:   :
Db	69	EKEKGDDGKKSEKGDGDKKEEEKDEBKDGDKEDDKDEKDEDKDEKDADEKDE	128			
		:   :	:   :	:   :	:   :	:   :
Qy	107	DLOREHSSQSSTDVTATVLDKNNISSKSTTNPNK	144			
		: : :   :	:   :	:   :	:   :	:   :
Db	129	EKKDDKDEKDKODEKKEKSKKSKSKSKSKSKS	166			
		: : :   :	:   :	:   :	:   :	:   :

RESULT 18

B72291  
hypothetical protein - Thermotoga maritima (strain MGB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: B72291  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richard  
C.M.

hypothetical protein BH4017 [imported] - *Bacillus halodurans* (strain C-125)  
C;Species: *Bacillus halodurans*

C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
C/Accession: A84152  
R/Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiraoka, Y. Nucleic Acids Res. 28, 4317-4331, 2000  
A/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A/Reference number: A83650; MUID:20512582; PMID:11058132  
A/Accession: A84152  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-614 <STO>  
A/Cross-references: GB:AP001520; GB:BA000004; NID:gl0176401; PIDN:BA07736.1; GSPDB:GN000000000  
A/Experimental source: strain C-125  
C/Genetics:  
A/Gene: BH4017

Query Match	12.1%	Score 90.5	DB 2	Length 614
Best Local Similarity	26.1%	Pred. No. 17		
Matches	37	Conservative 26	Mismatches 42	Indels 37
				Gaps 7

  

Qy	4	TVTIQNCKMSSTIVSE-----EDFTLPVYKGELEKGYQFDGW-----EISGPFEGKKDAG	53
Db	392	TYTLQTAIQM-TPIVNEVSPOTREEF-----ARKAHLDGWADVSKVTFYFAGRNIA	443
Qy	54	YVINLSKDTETIKPVFKKIEKK--EENKNPTFDVSKKK-----DNPQNVNHSQINSHRRKD	107
Db	444	QL-----GKIEEKLQDKYNNYTFDFPKKEVVNVQSPKSTSAIGKVGIVD	490
Qy	108	LQREHSQKSDSTKDVTATVLD	129
Db	491	LGRKYHGKEDLERLSKSQTE	512

**RESULT 23**

T28391 ORF MSV230 hypothetical protein - Melanoplus sanguinipes entomopoxvirus  
C&S:Species: Melanoplus sanguinipes entomopoxvirus  
C&Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C&Accession: T28391  
R&Afonso, C.B.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A&Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A&Reference number: Z20484; MUID:99102612; PMID:9847359  
A&Accession: T28391  
A&Status: preliminary; translated from GB/EMBL/DBJ  
A&Molecule type: DNA  
A&Residues: 1-670 <AFO>  
A&Cross-references: EMBL:AF063866; NID:G4049647; PIDN:AAC97713.1; PID:G4049  
C&Genetics:  
A&Note: MSV230

Query Match	12.1%	Score 90.5;	DB 2;	Length 670;
Best Local Similarity	22.5%;	Pred. No. 19;		
Matches 36;	Conservative 34;	Mismatches 69;	Indels 21;	Gaps 5;
Qy	5	VTIQNGKEMSTI-----VSEDFILPVYKGELEKG-----QFDGWELSGF	46	
Db	33	VNPEEKQIILSTLTKFNFDKTEMGVSVKVFQVLINNKSASEKYSVDVSSIDESQNSDS	92	
Qy	47	EGKKDAGVNLNSKDTIKPVFKKIE-EKKEEENKPTFDVSKKDNQVNHSQL-NESHR	104	
Db	93	DSDSDSGVNIDESQNSDSKVINKLENESQNSDSKVNIDESQNSDS-KVINKLENESQN	151	
Qy	105	KEDLQREHSHQSKSDTKDVTATVLDKNNISSKSTTNNPNK	144	
Db	152	SDSKVNIDESQNSDSKVNIDESQNSDSKVNIDESQNSDSK	191	

## RESULT 24

probable transcription factor SP7 - yeast (*Saccharomyces cerevisiae*)  
 N;Alternate names: protein YBR0739; protein YBR081C  
 C;Species: *Saccharomyces cerevisiae*  
 C;Date: 28-Jan-1994 #sequence revision 09-Sep-1994 #text change 20-Sep-1994

C; accession: S41552; S45946; S45948; S40800; S45478; S54985; S59716  
R; Gansheroff, L.; Dollard, C.; Tan, P.; Winston, F.  
submitted to the EMBL data library, July 1993  
A; reference number: S41552  
A; accession: S41552  
A; molecule type: DNA  
A; residues: 1-1332 <GAN>  
A; cross-references: EMBL:L22537; NID:G349189; PID:AAC37424.1; PID:G349190  
R; Steensma, H.Y.; van der Aart, Q.J.M.  
submitted to the Protein Sequence Database, August 1994  
A; reference number: S45932  
A; accession: S45946  
A; molecule type: DNA  
A; residues: 1-1332 <STB>  
A; cross-references: EMBL:X35950; NID:G536341; PID:CAA85026.1; PID:G536342  
R; Andre, B.; Czapleuch, C.; Hein, C.; Jauniaux, J.C.; Urrestazu, A.; Vissers  
submitted to the Protein Sequence Database, August 1994  
A; reference number: S45893

Query Match 12.1%; Score 90.5; DB 2; Length 1332;  
Best Local Similarity 24.5%; Pred. No. 41;  
Matches 39: Conservative 26; Mismatches 57; Indels 37

N;Alternate names: protein YBR0739; protein YBR081C  
C;Species: *Saccharomyces cerevisiae*  
C;Date: 28-Jan-1994 #sequence revision 09-Sep-1994 #text change 20-Sep-1999

```
QY 65 K--PVFKKIEKKEENKPTFD-----VSKKKO-----NPQVNHSQLNESH 103
Db 595 KENGKVSQSSKTVKDEAPTNDKLTLSVPEGEKEKDTASSTVTVHNVNKNEIKENG 654
QY 104 RKEDLQR-FEHQKSDSTKDVATVLD-KNNISSKSTTN 140
Db 655 KNEEQDMVEESKTEDSSKDAADAAKOTDGLQDKTAEN 693

RESULT 25
T28676
rhoptry protein - plasmodium yoelii (fragment)
C:Species: Plasmodium yoelii
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 01-Dec-2000
C:Accession: T28676; A45521
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.
Mol. Biochem. Parasitol. 76, 329-332, 1996
A:Title: Comparison of two members of a multigene family coding for high-molecular mass
A:Reference number: Z20507; MUID:97077455; PMID:8920022
A:Accession: T28676
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2401 <SIN>
A:Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.
Mol. Biochem. Parasitol. 42, 241-246, 1990
A:Title: Identification of the gene for a plasmodium yoelii rhoptry protein. Multiple co
A:Reference number: A45521; MUID:91101660; PMID:2270106
A:Accession: A45521
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 2260-2401 <KE>
A:Cross-references: GB:M34281

Query Match 12.1%; Score 90.5; DB 2; Length 2401;
Best Local Similarity 27.4%; Pred. No. 79;
Matches 43; Conservative 26; Mismatches 53; Indels 35; Gaps 6;

QY 6 TTONGEMSSSTI-----VSEDFILPVYKGEKGY---QPDGWEISGFEGKKDA 52
Db 924 TKNHEKISDRKNSLKIIOFSEESYINDI-KKELEKNVLESQNNNTDINQYLSKIEN 982
QY 53 GYVINLSKDTFKPVFKKIEKKEENKPTFDVSKKKONPQVNHSQLNESHKRLQREE 112
Db 983 IV--NLKLNKIKIIDKVKYETDETEK-----NNKKINAELSNSEKIITQLKENSLLKE 1035
QY 113 HSQKSDSTKD-----VTATVLDKNNISS 135
Db 1036 COSKIKSTIDDNVSECIKNTLNKTYIVNKKNNINT 1072

RESULT 26
B81594
hypothetical protein CP0281 [imported] - Chlamydomophila pneumoniae (strain AR39)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C:Accession: B81594
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255; PMID:10684935
A:Accession: B81594
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <REA>
A:Cross-references: GB:AE002161; NID:g7189205; PIDN:AAF38139.1; PID:g718920
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0281

Query Match 12.0%; Score 89.5; DB 2; Length 508;
```

```
Best Local Similarity 24.5%; Pred. No. 16;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 64 IKPVFKKIEKKEENKPTFD-----VSKKKONPQVNHSQLNESHK 105
Db 92 VKGVFKKTPQARPEVSSPRLPSHVHGQRLPGLEGFRDRIQKRSNPADLGKMKRSYSD 151
QY 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139
Db 152 GDLDRVGHDSNEDSTEDSRS---EGGEPSSKSS 182

RESULT 27
C72074
hypothetical protein - Chlamydomophila pneumoniae (strain CW1029)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C:Accession: C72074
R:Kelman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: C72074
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <ARN>
A:Cross-references: GB:AE001632; GB:AE001363; NID:g4376755; PIDN:AAD18613.1; PID:g437675
A:Experimental source: strain CW1029
C:Genetics:
A:Gene: CPn0473

Query Match 12.0%; Score 89.5; DB 2; Length 508;
Best Local Similarity 24.5%; Pred. No. 16;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 64 IKPVFKKIEKKEENKPTFD-----VSKKKONPQVNHSQLNESHK 105
Db 92 VKGVFKKTPQARPEVSSPRLPSHVHGQRLPGLEGFRDRIQKRSNPADLGKMKRSYSD 151
QY 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139
Db 152 GDLDRVGHDSNEDSTEDSRS---EGGEPSSKSS 182

RESULT 28
E86549
hypothetical protein CPJ0473 [imported] - Chlamydomophila pneumoniae (strain J138)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: E86549
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; I
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10871362
A:Accession: E86549
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-508 <STO>
A:Cross-references: GB:BA000008; NID:g8978843; PIDN:BAA98679.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CPJ0473

Query Match 12.0%; Score 89.5; DB 2; Length 508;
Best Local Similarity 24.5%; Pred. No. 16;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 64 IKPVFKKIEKKEENKPTFD-----VSKKKONPQVNHSQLNESHK 105
Db 92 VKGVFKKTPQARPEVSSPRLPSHVHGQRLPGLEGFRDRIQKRSNPADLGKMKRSYSD 151
QY 106 EDLQREHSQKSDSTKDVATVLDKNNISSKSTT 139
```

Db 152 GDLDRVGHSDNEDSTEDRS---EGGEPSSKSS 182

RESULT 29

S67610

probable membrane protein YDL074c - yeast (Saccharomyces cerevisiae)

N;Alternate names: hypothetical protein D2483

C;Species: Saccharomyces cerevisiae

C;Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002

C;Accession: S67610

R;Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.

submitted to the Protein Sequence Database, July 1996

A;Reference number: S67608

A;Accession: S67610

A;Molecule type: DNA

A;Residues: 1-700 <WAM>

A;Cross-references: EMBL:Z74122; NID:gl431087; PID:e253213; PID:gl431088; GSPDB:GN000004;

A;Experimental source: strain S288C

C;Genetics:

A;Gene: SGD:BRE1; MIPS:YDL074c

A;Cross-references: SGD:S0002232

A;Map position: 4L

C;Keywords: transmembrane protein

F;69-85/Domain: transmembrane #status predicted <TM>

Query Match 12.0%; Score 89.5; DB 2; Length 700;

Best Local Similarity 23.5%; Pred. No. 24;

Matches 36; Conservative 29; Mismatches 47; Indels 41; Gaps 7;

Qy 9 NGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGK----- 50

Db 430 NDTKSSST---QDALIKEIQ-DLEKGRP---ELSLTHKKYSEIINHESVSKLTVEK 480

Qy 51 ---DAGYVNL-SKDTF---IKPVFKLIEKKEENKPTFDVSKKKNQPNVNHSLNESH 103

Db 481 TKADQKYFAAWRSKDSILIEIKTSLSKSNEL-----ILQKSDRLLOQKIGMLH 533

Qy 104 RKEDLREHSQKSDSTVDVTATVLDKNNISSK 136

Db 534 KQLDLSQNNERRLDSSKTEIKLIDLNNTSTK 566

RESULT 30

T19006

ankyrin related protein C06C3.1 - Caenorhabditis elegans

N;Contains: myosin-light-chain-phosphatase (EC 3.1.3.53)

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Mar-2000

C;Accession: T19006; T22086

R;Berks, M.

submitted to the EMBL Data Library, August 1994

A;Reference number: Z19058

A;Accession: T19006

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1016 <WIL>

A;Cross-references: EMBL:Z36719; PIDN:CAA85318.1; GSPDB:GN000020; CESP:C06C3.1

A;Experimental source: clone C06C3

R;Matthews, P.

submitted to the EMBL Data Library, January 1995

A;Reference number: Z19510

A;Accession: T22086

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1016 <WT2>

A;Cross-references: EMBL:Z47809; PIDN:CAA87782.1; GSPDB:GN000020; CESP:C06C3.1

A;Experimental source: clone F42A8

C;Genetics:

A;Gene: CESP:C06C3.1

A;Map position: 2

A;Introns: 27/3; 94/3; 279/3; 352/2; 462/2; 523/3; 569/2; 718/3; 766/3; 833/3; 89

C;Keywords: phosphoric monoester hydrolase

Query Match 12.0%; Score 89.5; DB 2; Length 1016;

Best Local Similarity 21.6%; Pred. No. 36;

Matches 33; Conservative 33; Mismatches 38; Indels 49; Gaps 8;

Qy 5 VTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDGAGYVNLISKDTFI 64

Db 265 ILLENGAELSD-----LTTGADVLGVADKECIDYVELA-DTV- 302

Qy 65 KPVFKLIEKKEE---NKPTFDVSKKKNQPNVNHQ-LNESHKREDIQ-BEEHSQK--- 116

Db 303 ----KVQNRKSPGSGSQPPTSILQEKVHRMPSHEEHVLTSEKRDLOHQDQHSNEFL 357

Qy 117 -----SDSTKDVYAT-VLDKNNISS 135

Db 358 HSHPSTASVGTSTSSNTNTTTTIVIGENDISA 390

RESULT 31

E96795

unknown protein F28016.8 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: E96795

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: E96795

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-528 <STO>

A;Cross-references: GB:AB005173; NID:G6143888; PIDN:AAF04434.1; GSPDB:GN00141

C;Genetics:

A;Gene: F28016.8

A;Map position: 1

Query Match 11.9%; Score 89; DB 2; Length 528;

Best Local Similarity 24.3%; Pred. No. 19;

Matches 35; Conservative 22; Mismatches 51; Indels 36; Gaps 6;

Qy 31 ELEKGYQFDGW-----EISGPEGKDDAG-----YVNLKDTFIKPVFKKIEE 73

Db 200 ELAYDYNFEWYGGAKVRCLCGAVACSGFLGAKSRGFQASPYVILIEDTY---VVEDGDD 256

Qy 74 KKEENKPTFDVSKK--DNPQVNHSLNESHKREDLQREH-----SQKSPDKD- 122

Db 257 RYSVDKIPVYDSADELTSEPSKNGESNTNEEKDKISTENHLESTALNIQQSDSTPT 316

Qy 123 -----VTATVLDKNNISSKSTNN 141

Db 317 MEEDVVTETVKTETSEDMKLLSQN 340

RESULT 32

E89883

conserved hypothetical protein SA0976 [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C;Accession: E89883

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc,

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: E89883





Db 222 SNRRDENTIEDES 236

RESULT 40

A41859

IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus C;Species: Haemophilus influenzae A;Variety: strain HK715

Search completed: February 10, 2004, 10:58:33  
Job time : 12.3976 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 32.7596 Seconds  
(without alignments)  
697.707 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVTVTIQNGKMSSTIVSE.....ATVLKKNISKSTNNPNK 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

```

1: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqp-embl/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	748	100.0	773	22 AAB48343	S. pneumoniae Sp13
2	748	100.0	2120	21 AAY81710	Streptococcus pneu
3	748	100.0	2140	24 ABU01020	S. pneumoniae type
4	615	82.2	117	19 AAW55096	Streptococcus pneu
5	615	82.2	117	23 ABP54590	S. pneumoniae Sp04
6	110	14.7	746	22 AAG81779	S. epidermidis ope
7	110	14.7	778	23 ABP39023	Staphylococcus epi
8	106.5	14.2	484	21 AAG47777	Arabidopsis thalia
9	104	13.9	665	21 AAB18278	Plasmodium falcipa

10	101.5	13.6	564	22 ABB61977	Drosophila melanog
11	92.5	12.4	2519	22 ABG16636	Novel human diagno
12	91	12.2	2060	23 AAE20967	Staphylococcus lug
13	89.5	12.0	511	20 AAY35091	Chlamydia pneumoni
14	89.5	12.0	1408	22 ABB58704	Drosophila melanog
15	89	11.9	645	24 ABJ19106	Pathogen specific
16	89	11.9	654	24 ABP56879	Staphylococcus epi
17	88.5	11.8	281	22 ABB64828	Drosophila melanog
18	88.5	11.8	1141	22 AAG85008	Shrimp white spot
19	88	11.8	225	23 ABP73992	Candida albicans e
20	88	11.8	258	22 AAB94584	Human protein sequ
21	88	11.8	645	24 ABJ18979	Pathogen specific
22	88	11.8	817	22 AAM79318	Human protein SEQ
23	88	11.8	817	22 AAM79319	Human protein SEQ
24	87	11.6	607	22 AAM39950	Human polypeptide
25	87	11.6	635	21 AAB18176	Plasmodium falcipa
26	87	11.6	902	23 ABP40312	Staphylococcus epi
27	86.5	11.6	1183	22 ABB58769	Drosophila melanog
28	86.5	11.6	2485	21 AAB18172	Plasmodium falcipa
29	86	11.5	209	21 AAB42897	Human ORFX ORF2661
30	86	11.5	408	21 AAG37134	Arabidopsis thalia
31	86	11.5	456	21 AAG37133	Arabidopsis thalia
32	86	11.5	476	21 AAG37132	Arabidopsis thalia
33	86	11.5	757	21 AAY44384	Human cell cycle r
34	86	11.5	758	21 AAB53319	Human colon cancer
35	86	11.5	922	22 AAG85023	Shrimp white spot
36	86	11.5	1315	20 AAY08642	S. aureus SdrD pro
37	86	11.5	1315	24 ABJ18969	Pathogen specific
38	86	11.5	1349	22 AAU34402	Staphylococcus aur
39	86	11.5	1349	22 AAU37544	Staphylococcus aur
40	86	11.5	1980	21 AAB23563	Human sodium chann
41	86	11.5	1980	23 AAO14927	Human sodium chann
42	85.5	11.4	402	20 AAY03190	S. aureus trigger
43	85.5	11.4	434	20 AAY08215	Staphylococcus hae
44	85.5	11.4	525	20 AAY03189	S. aureus trigger
45	85.5	11.4	884	23 ABP73375	Candida albicans e

#### ALIGNMENTS

#### RESULT 1

AAAB48343  
ID AAB48343 standard; Protein; 773 AA.

XX AC AAB48343;

XX AC AAB48343;

DT 20-APR-2001 (first entry)

XX S. pneumoniae Sp130 polypeptide.

DE Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;  
KW bronchial; lung; blood; infection; immune response; immunotherapy;  
KW antibacterial; auditory; vaccine.

XX Streptococcus pneumoniae.

OS Streptococcus pneumoniae.

PN WO200076540-A2.

XX 21-DEC-2000.

PF 09-JUN-2000; 2000WO-US15925.

XX 10-JUN-1999; 99US-0138453.

XX (MEDI-) MED IMMUNE INC.

PI Adamou JE, Choi GH;

XX WPI; 2001-112197/12.

DR N-PSDB; AAC84742.

XX New vaccines comprising Sp128 or Sp130 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by  
 PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
 PT blood infections  
 XX  
 XX  
 XX Claim 8; Page 51-54; 54pp; English.  
 PS  
 CC The invention relates to novel immunogenic polypeptides, Sp128 and Sp130  
 CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
 CC the treatment and prevention of pneumococcal infections, particularly  
 CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
 CC bronchial, lung or blood infections. The antigens are used as immunogenic  
 CC agents to stimulate an immune response. The antisera and antibodies may  
 CC also be used in diagnosing and treating pneumococcal infections.  
 CC Recombinant polypeptides serve as a mechanism for stimulating production  
 CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
 CC as reagents in other processes such as affinity chromatography. The  
 CC present sequence represents the S. pneumoniae Sp130 polypeptide.  
 XX  
 XX  
 SQ Sequence 773 AA;

Query Match 100.0%; Score 748; DB 22; Length 773;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-67;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HRTVTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSK 60  
 DB 630 HRTVTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSK 689  
 QY 61 DTFIKPVFKIIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 120  
 DB 690 DTFIKPVFKIIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 749  
 QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
 DB 750 KDVTATVLDKNNISSKSTNNPNK 773

RESULT 2  
 AAY81710  
 ID AAY81710 standard; Protein; 2120 AA.  
 XX  
 AC AAY81710;  
 XX  
 DT 02-JUN-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae protein sequence ID3.  
 XX  
 KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
 KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
 KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200006738-A2.  
 XX  
 PD 10-FEB-2000.  
 XX  
 XX 27-JUL-1999; 99WO-GB02452.  
 XX  
 XX 27-JUL-1998; 98GB-0016336.  
 PR 19-MAR-1999; 99US-0125329.  
 XX  
 XX (MICR-) MICROBIAL TECHNIQS LTD.  
 XX  
 XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
 XX  
 DR WPI; 2000-195301/17.  
 DR N-PSDB; AAZ91806.  
 XX  
 XX Streptococcal proteins and polynucleotides useful for diagnosis,  
 PT treatment and prophylaxis of bacterial infections  
 PT  
 PT

PS Claim 2; Page 41-42; 76pp; English.  
 XX  
 CC This sequence represents a Streptococcus pneumoniae protein of the  
 CC invention. The proteins (or their homologues, derivatives and/or  
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
 CC compositions comprising the proteins are useful as vaccines and also in  
 CC diagnostic assays. The sequences are useful for the detection or  
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
 CC with them. Agents capable of antagonising, inhibiting or interfering with  
 CC the function or expression of the protein or polypeptide are useful in  
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
 CC infection. As the sequences can be used to treat S. pneumoniae infection,  
 CC they can be used to treat bacterial pneumonia, which has high rates in  
 CC young children, the elderly, and in patients with predisposing conditions  
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
 CC or with immunosuppressive disorders, especially AIDS. They can also be  
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
 CC meningitis.  
 XX  
 XX  
 SQ Sequence 2120 AA;

Query Match 100.0%; Score 748; DB 21; Length 2120;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-67;  
 Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HRTVTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSK 60  
 DB 1943 HRTVTIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSK 2002  
 QY 61 DTFIKPVFKIIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 120  
 DB 2003 DTFIKPVFKIIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDST 2062  
 QY 121 KDVTATVLDKNNISSKSTNNPNK 144  
 DB 2063 KDVTATVLDKNNISSKSTNNPNK 2086

RESULT 3  
 ABU01020  
 ID ABU01020 standard; Protein; 2140 AA.  
 XX  
 AC ABU01020;  
 XX  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain protein from coding region #590.  
 XX  
 KW Bacterial meningitis; pneumonia; sepsis; otitis media;  
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KW auditory; respiratory; gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae type 4 strain.  
 XX  
 PN WO200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 XX 27-MAR-2002; 2002WO-IB02163.  
 XX  
 XX 27-MAR-2001; 2001GB-0007658.  
 PR  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 XX Masignani V, Tettelin H, Fraser C;  
 XX  
 DR WPI; 2003-040579/03.  
 DR N-PSDB; ABX06302.  
 XX  
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
 PT

or ear infection -

Claim 1; SEQ ID No 1180; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the Streptococcus pneumoniae type 4 strain genomic sequence appearing as AB556454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a Streptococcus nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a Streptococcus pneumoniae bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to Streptococcus bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2469 proteins expressed by the identified coding regions from the genomic sequence.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 2140 AA;

Query Match 100.0%; Score 748; DB 24; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 7.1e-67;  
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGVYINLSK 60  
Db 1963 HRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGVYINLSK 2022

Qy 61 DTFIKPVFKIIEKKKEENKPTFDVSKKKDNPQVNHSQLNESHKEDLQREHSQKSDST 120  
Db 2023 DTFIKPVFKIIEKKKEENKPTFDVSKKKDNPQVNHSQLNESHKEDLQREHSQKSDST 2082

Qy 121 KDVTATVLDKNNISSKSTNNPNK 144  
Db 2083 KDVTATVLDKNNISSKSTNNPNK 2106

RESULT 4

AAW55096

ID AAW55096 standard; Protein; 117 AA.

XX

AC AAW55096;

XX

DT 02-OCT-1998 (first entry)

DE Streptococcus pneumoniae SP0043 protein.

XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
KW detection; pneumonia; otitis media; meningitis.

XX Streptococcus pneumoniae.

XX WO9818930-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US19422.  
XX 31-OCT-1996; 96US-0029960.  
PR (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;  
XX  
DR WPI; 1998-272224/24.  
DR N-PSDB; AAV27357.

Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae - or their epitope-containing fragments, useful in protective or therapeutic vaccines, and for diagnosis

PS Claim 11; Page 62; 118pp; English.

XX The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.

XX Sequence 117 AA;

Query Match 82.2%; Score 615; DB 19; Length 117;  
Best Local Similarity 100.0%; Pred. No. 5e-55;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 YKGELEKGYQFDGWEISGFEGKKDAGVYINLSKDTFKPVFKIIEKKKEENKPTFDVSK 87  
Db 1 YKGELEKGYQFDGWEISGFEGKKDAGVYINLSKDTFKPVFKIIEKKKEENKPTFDVSK 60

Qy 88 KCDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTNNPNK 144

Db 61 KCDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTATVLDKNNISSKSTNNPNK 117

RESULT 5

ABP54590

ID ABP54590 standard; Protein; 117 AA.

XX

AC ABP54590;

XX

DT 04-SEP-2002 (first entry)

DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
KW antibacterial; Streptococcal infection; detection.

XX Streptococcus pneumoniae.

XX US2002061545-A1.

XX 23-MAY-2002.

XX 22-JAN-2001; 2001US-0765272.

XX 30-OCT-1997; 97US-0961083.

XX (CHOI/) CHOI G H.

XX (KUNS/) KUNSCH C A.

PA (BARA/) BARASH S C.  
 PA (DILL/) DILLON P J.  
 PA (DOUG/) DOUGHERTY B.  
 PA (FANN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 XX  
 XX WPI; 2002-479261/51.  
 DR N-PSDB; ABQ84825.  
 XX  
 XX New Streptococcus pneumoniae antigens, useful for detecting  
 PT Streptococcus and for preventing or attenuating disease caused by  
 PT Streptococcus infection -  
 XX  
 PS Claim 11; Page 29; 70pp; English.  
 XX  
 XX ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
 CC The S. pneumoniae antigens have antibacterial activity and can be  
 CC used in vaccines. The S. pneumoniae antigens can also be used to  
 CC prevent or attenuate a streptococcal infection in an animal. The  
 CC polynucleotides encoding the S. pneumoniae antigens can be used to  
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
 CC which are used in an example from the present invention.  
 XX  
 SQ Sequence 117 AA;  
 Query Match 82.2%; Score 615; DB 23; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 5e-55;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 28 YKGELEKGYQFDGWETSGEGKDDAGYVNLKDTFIKPVFKKIEEKEENKPTFDVSK 87  
 Db 1 YKGELEKGYQFDGWETSGEGKDDAGYVNLKDTFIKPVFKKIEEKEENKPTFDVSK 60  
 QY 88 KKDNPQVNSHLSHREKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 144  
 Db 61 KKDNPQVNSHLSHREKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117  
 RESULT 6  
 AAG81779  
 ID AAG81779 standard; Protein; 746 AA.  
 XX  
 AC AAG81779;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 XX S. epidermidis open reading frame protein sequence SEQ ID NO:652.  
 DE  
 XX Staphylococcus epidermidis SR1 strain; infection; diagnosis;  
 KW vaccination; endocarditis.  
 XX  
 XX Staphylococcus epidermidis.  
 XX  
 XX WO200134809-A2.  
 PN  
 XX  
 XX 17-MAY-2001.  
 PD  
 XX  
 XX 09-NOV-2000; 2000WO-US30782.  
 PF  
 XX  
 XX 09-NOV-1999; 99US-0164258.  
 PR  
 XX  
 XX (GLAX ) GLAXO GROUP LTD.  
 PA  
 XX  
 XX Kimmerly WJ;  
 PI  
 XX  
 XX WPI; 2001-316495/33.  
 DR  
 XX  
 DR N-PSDB; AAG82629.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 PT useful for vaccinating against infections, e.g. endocarditis -  
 XX  
 XX Claim 18; Page 208; 218pp; English.  
 XX  
 XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX  
 SQ Sequence 746 AA;  
 Query Match 14.7%; Score 110; DB 22; Length 746;  
 Best Local Similarity 27.0%; Pred. No. 0.027;  
 Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;  
 QY 5 VTIQNGKMSSTIVSEEDFILPVYK-----GELEKGYQFDGW-----EISGFE-----G 48  
 Db 596 ITIGNGKQIKQGSVKSGTKVLPHSKVLMLTGELTMP-DMTGWTKEDVLAFEDLTKLKVS 654  
 QY 49 KKDAGYVNL--LSKDTFIKPVFKKIEEKEENKPTFDVS-----KKDNPQVNSHLSHRE 102  
 Db 655 TKGNGFVTVNQISKGQIIK-----NKKDIEVLSAEDTDDQKTDDESDN 701  
 QY 103 HRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 141  
 Db 702 KSKDKADEHSHNTSSSTKN-----DKSNADSKNDSDD 734  
 RESULT 7  
 ABP39023  
 ID ABP39023 standard; Protein; 778 AA.  
 XX  
 AC ABP39023;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.  
 DE  
 XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.  
 XX  
 XX Staphylococcus epidermidis.  
 OS  
 XX  
 XX US6380370-B1.  
 PN  
 XX  
 XX 30-APR-2002.  
 PD  
 XX  
 XX 13-AUG-1998; 98US-0134001.  
 PF  
 XX  
 XX 14-AUG-1997; 97US-055779P.  
 PR  
 XX  
 XX 08-NOV-1997; 97US-064964P.  
 PR  
 XX  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA  
 XX  
 XX Doucette-Stamm LA, Bush D;  
 PI  
 XX









PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 XX biodiversity -  
 XX  
 XX Claim 20; SEQ ID No 46995; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (II) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: the sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2519 AA;  
 Query Match 12.4%; Score 92.5; DB 22; Length 2519;  
 Best Local Similarity 31.0%; Pred. No. 8.8;  
 Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 58 LSKDTFIKPVFKIEEENKPTFDVSKKDNQVNHSQLNESHKEDLORE-----E 112  
 Db 689 VKETFKVP-----EDKKEKPKKEVAKEDKTPF----KKEKPKKEEVKKEVKEIK 740

QY 113 HSQKSDSTKDV 123  
 Db 741 KEEKPEKKEV 751

RESULT 12  
 AAE20967  
 ID AAE20967 standard; Protein; 2060 AA.  
 XX  
 AC AAE20967;  
 XX  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus lugdunensis von Willebrand factor binding protein.  
 XX  
 KW Von Willebrand factor binding protein; vwb; immunogen; antibacterial;  
 KW vaccine; infection.  
 XX  
 OS Staphylococcus lugdunensis.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..47  
 FT /label= Signal\_peptide  
 FT Protein 48..2060  
 FT /note= "Mature von Willebrand factor binding protein"  
 XX  
 PN WO200228892-A1.  
 XX  
 PD 11-APR-2002.

XX  
 XX 06-APR-2001; 2001WO-SE00766.  
 XX  
 PR 04-OCT-2000; 2000SE-0003573.  
 XX  
 PA (BIOS-) BIOTAPRO AB.

XX Guss B, Frykberg L, Jacobsson K, Ahlen J, Nilsson M;  
 PI WPI; 2002-304928/34.  
 DR N-PSDB; AAD33371.  
 XX  
 XX New von Willebrand factor binding protein from Staphylococci, useful  
 PT for determining and treating staphylococcal infection -  
 XX  
 PS Claim 3; Page 35-41; 53pp; English.  
 XX  
 CC The present invention relates to von Willebrand factor binding protein or  
 CC polypeptide (vwb) from Staphylococci. The vwb and immunogens of vwb are  
 CC useful in vaccines to combat infections caused by Staphylococci. The  
 CC invention is also useful for detection of staphylococcal infection and  
 CC purifying von Willebrand factor from a complex solution. The present  
 CC sequence is Staphylococcus lugdunensis vwb1 protein.

XX Sequence 2060 AA;

Query Match 12.2%; Score 91; DB 23; Length 2060;  
 Best Local Similarity 25.0%; Pred. No. 9.5;  
 Matches 36; Conservative 24; Mismatches 48; Indels 36; Gaps 8;

QY 2 RVTVTIQ-NGKEMSTIVSEEDFILPVYKGELEKGYQFDG--WEISGFEKGDAGYVINL 58  
 Db 1919 KITTYVDVTNGRE-----IVPSRKGLPP-EQFIGQDMQYTGHK-----I 1956

QY 59 SKDTFIKPVFKIEE-----KKEEENKPTFDVSKKDNQVNHSQLNESHKKE--DLQ 109  
 Db 1957 EKDGTITTYIKKVENAVPAKQLKTKYN--TQSESOQKHTFQVKQOLVKYHNKVEORSIE 2014  
 QY 110 REHSQKSDSTKDVATVLDKNNI 133  
 Db 2015 KSEHTDMHVSELPETGETANKGL 2038

RESULT 13

AAV35091  
 ID AAV35091 standard; Protein; 511 AA.

AC AAV35091;

XX 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae transmembrane protein sequence.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN WO927105-A2.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

XX (GEST ) GENSET.

PI Griffais R;

DR WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX Page 975-976; Disclosure; 1912pp; English.

XX AAY34584-Y35879 represent the proteins encoded by all the open reading

[illegible]

RESULT 15	
ABJ19106	
ID	ABJ19106 standard; Protein; 645 AA.
XX	
XX	ABJ19106;
XX	
XX	06-MAR-2003 (first entry)
DT	
XX	
DE	Pathogen specific antigen related staphylococcal protein SEQ ID No 414.
XX	
KW	Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;
KW	hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;
KW	auto-immunity; vaccine; staphylococcal infection; antibody; cancer;
KW	autoimmune disease; HIV; hepatitis.
XX	
OS	Staphylococcus sp.
XX	
XX	WO200259148-A2.
PN	
XX	
PD	01-AUG-2002.
XX	
PF	21-JAN-2002; 2002WO-EP00546.
XX	
PR	26-JAN-2001; 2001AT-0000130.
XX	
PA	(CIST-) CISTEM BIOTECHNOLOGIES GMBH.
XX	
PI	Meinke A, Nagy E, Von Ahnen U, Klade C, Henics T, Zauner W;
PI	Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;
PI	Tempelmaier B;
XX	
DR	WPI; 2003-075410/07.
XX	
PT	Identifying, isolating and producing hyperimmune serum-reactive
PT	antigens from a pathogen, for preparing vaccine or medicament for
PT	treating or preventing e.g. staphylococcal infections, comprises
PT	providing antibody preparation -
XX	
PS	Example 7; Page 220; 252pp; English.
XX	
CC	The invention relates to a novel method for identifying, isolating and
CC	producing hyperimmune serum-reactive antigens from a pathogen, tumour,
CC	allergen, a tissue or host prone to auto-immunity, where the antigens
CC	are used in a vaccine, comprises providing antibody preparation from a
CC	plasma pool of a type of animal, or individual sera with antibodies
CC	against the specific pathogen, tumour, allergen, tissue or host prone to
CC	auto-immunity. The hyperimmune serum-reactive antigens comprising any of
CC	the 62 sequences of 53-2261 amino acids fully defined in the

CC specification, or their hyperimmune fragments are useful for the  
CC manufacture of a pharmaceutical preparation, particularly a vaccine  
CC against staphylococcal infections or colonisation against *S. aureus* or *S.*  
CC epidermidis. The preparation of antibodies is useful for the manufacture  
CC of a medicament for treating or preventing staphylococcal infections or  
CC colonisation against *S. aureus* or *S. epidermidis*. The antibody  
CC preparations may also be used for diagnostic and imaging purposes. Other  
CC conditions that can be treated include cancer, autoimmune diseases or  
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or  
CC protozoan pathogens. This sequence represents a staphylococcal protein  
CC relating to the method for identifying and producing pathogen specific  
CC antigens of the invention.

XX  
SQ Sequence 645 AA;

Query Match 11.9%; Score 89; DB 24; Length 645;  
Best Local Similarity 21.9%; Pred. No. 3;  
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;

QY 6 TIONCKEMSSITVSEEDFILPVYKGELE-KGYQF-----DGWEISGFEGKK----- 50  
DB 355 SVNNESMMDTFVKH-----PIKTGMLNGKKYMMETNDDYWKDFWVGQVRRTISKDA 409  
QY 51 -----DAGYVINL-SKDTFIKPVFKKIEKKEE 78  
DB 410 KNNTRTIIFPYVEGKTLDAIVKHVKTIDYDQYHVRIVDKAFTKANTDKSNKKEQD 469  
QY 79 NKPTPDV-----SKKKNPQVNHSQLNESHKEDLQ-----REHSQKSDSTKQVT-ATVL 128  
DB 470 NSAKKEATPATPSKPTSPFVEKESQKQDSQKDDNKQLPSVEKENDESSESGKDKTPATKP 529  
QY 129 DKNNISSKSTTNPNK 144  
DB 530 TKGEVSSSTT--PTK 543

## RESULT 16

ABP56879  
ID ABP56879 standard; Protein; 654 AA.

XX  
AC ABP56879;

XX  
DT 07-APR-2003 (first entry)

XX  
DE Staphylococcus epidermidis KrkN protein SEQ ID NO:10.

XX  
KW Staphylococcus aureus; Staphylococcus epidermidis; MSCRAMM; antibody;  
KW microbial surface component recognising adhesive matrix molecule;  
KW surface protein; infection; antibacterial; antiinflammatory; vaccine;  
KW immunosuppressive; antiarthritic; gene therapy; pneumonia; endocarditis;  
KW septic arthritis; biomaterial related infection.

XX  
OS Staphylococcus epidermidis.

XX  
PN WO2002102829-A2.

XX  
PD 27-DEC-2002.

XX  
PF 17-JUN-2002; 2002WO-US19220.

XX  
PR 15-JUN-2001; 2001US-298098P.

XX  
PA (INH-) INHIBITEX INC.  
PA (QUEE-) QUEEN ELIZABETH COLLEGE DUBLIN.  
PA (UYPA-) UNIV PAVIA.

XX  
PI Foster TJ, Roche F, Patti JM, Hutchins JT, Hall A, Domanski P;  
PI Patel P, Syribey P, Speciale P;

XX  
DR WPI; 2003-167481/16.

XX  
DR N-PSDB; AB222903.

XX  
PT New isolated antibody that binds to a staphylococcal surface protein,

PT useful for treating or preventing Staphylococcus aureus infections,  
PT such as pneumonia, septic arthritis, endocarditis or biomaterial  
PT related infections

XX  
PS Claim 1; Page 29-30; 122pp; English.

CC The present invention describes an isolated antibody (I) that binds to a  
CC staphylococcal surface protein selected from any of the 12 sequences of  
CC 354-2283 amino acids given in ABP56879 to ABP56886. Also described:  
CC (1) an isolated antiserum (II) comprising (I); (2) a diagnostic kit, (III)  
CC comprising (I) and means for detecting binding by (I); (3) a diagnosing  
CC (MI) an infection of Staphylococcus aureus by adding (I) to a sample  
CC suspected of being infected with the infection, and determining if  
CC antibodies have bound to the sample; (4) a pharmaceutical composition  
CC (IV) for treating or preventing an infection of *S. aureus* comprising (I),  
CC and a vehicle, carrier or excipient; (5) treating (M2) or preventing an  
CC infection of *S. aureus* by administering (I) to a human or animal patient;  
CC (6) producing (M3) an immunological response by administering to a human  
CC or animal an immunogenic amount of the isolated surface protein; (7) an  
CC isolated active fragment (V) from the A domain of the DsgA protein; and  
CC (8) a vaccine (VI) for treating or preventing an infection of *S. aureus*  
CC comprising the surface protein in an amount effective to elicit an immune  
CC response, and a vehicle, carrier or excipient. (I) has antiinflammatory,  
CC antibacterial, immunosuppressive and antiarthritic activities, and can be  
CC used in gene therapy. The antibody, composition and vaccine are useful  
CC for treating or preventing Staphylococcus aureus infection in a human or  
CC animal, such as pneumonia, septic arthritis, endocarditis or biomaterial  
CC related infections. The present sequence represents Staphylococcus  
CC epidermidis KrkN protein, which is used in the exemplification of  
CC the present invention.

XX  
SQ Sequence 654 AA;

Query Match 11.9%; Score 89; DB 24; Length 654;  
Best Local Similarity 21.9%; Pred. No. 3.1;  
Matches 43; Conservative 26; Mismatches 63; Indels 64; Gaps 9;

QY 6 TIONCKEMSSITVSEEDFILPVYKGELE-KGYQF-----DGWEISGFEGKK----- 50  
DB 364 SVNNESMMDTFVKH-----PIKTGMLNGKKYMMETNDDYWKDFWVGQVRRTISKDA 418  
QY 51 -----DAGYVINL-SKDTFIKPVFKKIEKKEE 78  
DB 419 KNNTRTIIFPYVEGKTLDAIVKHVKTIDYDQYHVRIVDKAFTKANTDKSNKKEQD 478  
QY 79 NKPTPDV-----SKKKNPQVNHSQLNESHKEDLQ-----REHSQKSDSTKQVT-ATVL 128  
DB 479 NSAKKEATPATPSKPTSPFVEKESQKQDSQKDDNKQLPSVEKENDESSESGKDKTPATKP 538  
QY 129 DKNNISSKSTTNPNK 144  
DB 539 TKGEVSSSTT--PTK 552

## RESULT 17

ABE64828  
ID ABE64828 standard; Protein; 281 AA.

XX  
AC ABE64828;

XX  
DT 26-MAR-2002 (first entry)

XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 21276.

XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

XX  
OS Drosophila melanogaster.

XX  
PN WO200171042-A2.

XX  
PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL08931.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 21276; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 281 AA;  
 Query Match 11.8%; Score 88.5; DB 22; Length 281;  
 Best Local Similarity 22.1%; Pred. No. 1.1;  
 Matches 31; Conservative 34; Mismatches 52; Indels 23; Gaps 5;  
 QY 9 NGKMSSTVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVI-----NLS 59  
 DB 137 NFEFTNMTTQDILKNIY-SLLDK-----DNEGAIITSKELGMVIRALGRQPNESIA 188  
 QY 60 KDTFKIPVFKKIEE-KKEEENKPTFDVSKKDNQPNVHSQLNE-----SHRKEDLQREH 113  
 DB 189 KEFCNVLTKRMHMDTKKEELRDAFRVFDKENNGYISTTELRAVFMALGKLEDDLEEM 248  
 QY 114 SOKSDSTKDVATVLDKNNI 133  
 DB 249 IREYLDQDNHINFESNNM 268  
 RESULT 18  
 AAG85008  
 ID AAG85008 standard; Protein; 1141 AA.  
 AC AAG85008;  
 XX  
 DT 11-SEP-2001 (first entry)  
 DE Shrimp white spot Bacilliform virus (WSBV) protein 99.  
 XX  
 KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;  
 KW antiviral agent; gene expression; antisense construct;  
 KW transgenic viral resistant shrimp.  
 XX  
 OS White spot syndrome virus.  
 XX  
 PN WO200138351-A2.  
 XX  
 PD 31-MAY-2001.  
 XX  
 PF 08-NOV-2000; 2000WO-US28888.  
 XX  
 PR 24-NOV-1999; 99CN-0124717.  
 XX

PA (PENY-) PE CORP NY.  
 PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.  
 PA (SINO-) SINOGENOMAX CO LTD.  
 XX  
 PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;  
 XX  
 DR WPI; 2001-355877/37.  
 DR N-PSDB; AAH62788.  
 XX  
 PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus  
 PT (WSBV), useful for producing viral polypeptides that can be used to  
 PT screen for agents that are useful for treating WSBV infection -  
 XX  
 PS Claim 1; Figure 3; 626pp; English.  
 XX  
 CC The invention provides the primary nucleotide sequence of the WSBV genome  
 CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and  
 CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences  
 CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid  
 CC molecules and proteins of the invention are useful for diagnosis and  
 CC monitoring viral infection, in screens for antiviral agents and for  
 CC monitoring viral gene expression or activity during a treatment regimen.  
 CC The nucleic acid molecules are also useful as antisense constructs to  
 CC control viral gene expression in infected cells and tissues and to create  
 CC transgenic viral resistant shrimp.  
 XX  
 SQ Sequence 1141 AA;  
 Query Match 11.8%; Score 88.5; DB 22; Length 1141;  
 Best Local Similarity 23.3%; Pred. No. 7.5;  
 Matches 30; Conservative 31; Mismatches 43; Indels 25; Gaps 4;  
 QY 15 STIVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVINLSKDTFKIPVFKKIE 72  
 DB 621 SNVEEER-----EEEQNEEEEEVEEVEEGSKEDDGA-----PAQEEME 662  
 QY 73-EKKEEENKPTFDVSKKDNQPNVHSQLNESHRKEDLQREHSQKSDTKDVATVLDKNN 132  
 DB 663 EEKEEQ-----QQPEESNGENQEEEQQQQQQPEREENKQADSDSDSSSSSS 717  
 QY 133 ISSKSTNN 141  
 DB 718 SSSSSSSSS 726  
 RESULT 19  
 ABP73992  
 ID ABP73992 standard; Protein; 225 AA.  
 XX  
 AC ABP73992;  
 XX  
 DT 30-JAN-2003 (first entry)  
 DE Candida albicans essential protein SEQ ID NO 7829.  
 XX  
 KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;  
 KW signal transduction; DNA replication; cell division; growth;  
 KW proliferation; Candida albicans; fungicide; antifungal.  
 XX  
 OS Candida albicans.  
 XX  
 PN WO200253728-A2.  
 XX  
 PD 11-JUL-2002.  
 XX  
 PF 26-DEC-2001; 2001WO-US49486.  
 XX  
 PR 29-DEC-2000; 2000US-259128P.  
 PR 20-FEB-2001; 2001US-0792024.  
 PR 22-AUG-2001; 2001US-314050P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX



PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
XX WPI; 2002-566694/60.  
DR N-PSDB; AB232542.  
XX  
PT Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele  
PT of a gene and placing other allele of the gene under conditional  
XX expression -  
PS Claim 44; SEQ ID NO 7829; 167pp + Sequence Listing; English.  
XX  
XX The invention relates to constructing (M1) a strain of diploid fungal  
CC cells in which both alleles of a gene are modified, comprising modifying  
CC one allele by insertion or replacement by a cassette having an  
CC expressible selectable marker and modifying other allele by  
CC recombination, of a promoter replacement fragment with a heterologous  
CC promoter, so that expression of the second allele is regulated by the  
CC promoter. (M1) is useful for constructing a strain of diploid fungal  
CC cells in which both alleles of a gene are modified. The diploid fungal  
CC cells having both alleles modified are useful for identifying a gene that  
CC is essential to the survival or growth of a fungus, a gene that  
CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
CC that contributes to the resistance of a diploid fungus to an antifungal  
CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
CC and for identifying a therapeutic agent for treatment of a mammalian  
CC disease. (M1) is useful for identifying a compound which modulates the  
CC activity of a gene product, preferably enzymatic activity, carbon  
CC compound catabolism, biosynthetic, transporter, transcriptional,  
CC translational, signal transduction, DNA replication and cell division  
CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
XX the European Patent Office.  
XX  
SQ Sequence 225 AA;  
Query Match 11.8%; Score 88; DB 23; Length 225;  
Best Local Similarity 28.0%; Pred. No. 0.89; Indels 12; Gaps 4;  
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;  
QY 36 YQFDGWEISGFEKKDAGYVNLKDTFIKVPFKK----IEKKKEENKPTFVSKK-- 89  
DB 80 YDDDDDEFEFGSSNGAAKELNLSQAIEKWKQRDLIEIEKELNKKKEIEIEKAKS 139  
QY 90 --DNPQVNSQLNESHKEDLQREEH--SQKSDSTKDVDTATVLDKNN 132  
DB 140 TIDDFYENYSNRDNDHOKIELSEQEKFIKRDDFLK--RGTLDWRVN 184  
RESULT 20  
AAB94584  
ID AAB94584 standard; Protein; 258 AA.  
XX  
XX AAB94584;  
XX  
XX 26-JUN-2001 (first entry)  
XX  
XX Human protein sequence SEQ ID NO:15383.  
DE  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX  
XX EP1074617-A2.  
PN  
XX 07-FEB-2001.  
PD  
XX 28-JUL-2000; 2000EP-0116126.  
XX

PR 29-JUL-1999; 99JP-0248036.  
PR 27-AUG-1999; 99JP-0300253.  
PR 11-JAN-2000; 2000JP-0118776.  
PR 02-MAY-2000; 2000JP-0183767.  
PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 15383; 2537pp + CD ROM; English.  
PS  
XX The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 258 AA;  
Query Match 11.8%; Score 88; DB 22; Length 258;  
Best Local Similarity 26.3%; Pred. No. 1.1;  
Matches 31; Conservative 25; Mismatches 50; Indels 12; Gaps 4;  
QY 28 YKGELEKGYQFDGWEISGFEKKDAGY--VNLKDTFIKVPFKKIEKKKEENKPTFDV 85  
DB 39 WEGEDEDEDVKNWDDDDDDKEEAEVPEVKISEK---KKTAEKIEKEKQKKQKEEI 95  
QY 86 SKKKNPQ-----VNHSQLNESHKEDLQREHSHKSDSTKDVDTATV--LDKNNISSK 136  
DB 96 KKRLEPEPEPKVLTPEEQADKLRLKQLQESDLELAKETFGVNTVYIDAMNPSR 153  
RESULT 21  
ABJ18979  
ID ABJ18979 standard; Protein; 645 AA.  
XX  
XX ABJ18979;  
XX  
XX 06-MAR-2003 (first entry)  
XX  
XX Pathogen specific antigen related staphylococcal protein SEQ ID No 142.  
DE  
XX Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;  
KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;  
KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
KW autoimmune disease; HIV; hepatitis.  
XX



Db 315 T 315

RESULT 23  
AAM79319

ID AAM79319 standard; Protein; 817 AA.  
XX  
AC AAM79319;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human protein SEQ ID NO 2965.  
XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX

PN WO200157190-A2.  
XX

PD 09-AUG-2001.  
XX

PF 05-FEB-2001; 2001WO-US04098.  
XX

PR 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX

PA (HYSE-) HYSEQ INC.  
XX

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
XX

DR WPI; 2001-476283/51.  
DR N-PSDB; AAK52452.  
XX

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
XX useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 215; 5221pp; English.  
XX

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX

SQ Sequence 817 AA;  
XX

Query Match 11.8%; Score 88; DB 22; Length 817;  
Best Local Similarity 27.3%; Pred. No. 5.3;  
Matches 33; Conservative 21; Mismatches 55; Indels 12; Gaps 4;  
XX

QY 30 GELKGYQDGEWISG--FEGKDGAVINLSKDTFKVFKKIEKKEENKPTPDVSK 87  
DB 196 GQEKQESFSAEASGHQEVSPAVSLQKQDTSKLRSSTLPEQKKQIEKSPSPSQ 255

QY 88 -KKNPQVNHSLNESHKED-----LQREHSQKSDSTKYATATVLDKNNISSKST 138  
DB 256 WKQDTFKSAGYVQBEHKKQETPKLWVPVLOKEQ-DPKKQTPKSWTPSQSQNTTKSMT 314

QY 139 T 139  
DB 315 T 315

RESULT 24  
AAM39950

ID AAM39950 standard; Protein; 607 AA.  
XX  
AC AAM39950;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 3095.  
XX

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX

OS Homo sapiens.  
XX

PN WO200153312-A1.  
XX

PD 26-JUL-2001.  
XX

PF 26-DEC-2000; 2000WO-US34263.  
XX

PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX

PA (HYSE-) HYSEQ INC.  
XX

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX

DR WPI; 2001-442253/47.  
DR N-PSDB; AAI59106.  
XX

XX Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -  
XX  
XX Example 4; SEQ ID NO 3095; 10078pp; English.  
XX

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM39642-AAK42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders  
CC

Note: The sequence data for this patent did not form part of the printed

```
CC specification.
XX
SQ Sequence 607 AA;
Query Match 11.6%; Score 87; DB 22; Length 607;
Best Local Similarity 33.3%; Pred. No. 4.5;
Matches 20; Conservative 12; Mismatches 28; Indels 0; Gaps 0;
QY 59 SKDTFIKPVFKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLOREHKSQSD 118
Db 373 SPDNCCNELFKKKKKKEEKKKKKEEKKRRRRRREKREKREKREKREKREKD 432
RESULT 25
ID AAB18176 standard; Protein; 635 AA.
XX
AC AAB18176;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:33.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
FN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
XX (CARU/) CARUCCI D.
XX (GARD/) GARDNER M.
XX (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX
DR
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 85-86; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially when they are rifins or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasitic lifecycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
```

---

```
SQ Sequence 635 AA;
Query Match 11.6%; Score 87; DB 21; Length 635;
Best Local Similarity 23.8%; Pred. No. 4.8;
Matches 30; Conservative 20; Mismatches 52; Indels 24; Gaps 4;
QY 32 LEKGYQFDGWEISGFEGKKGAGYVINLSKDTFFIKPVFKIEKKEENKPTFDVSKKDN 91
Db 202 LQKYNIQDDDEEDNETIRSDSKLDRIYDSQSKDIMSSSPNKEES-----NSSDNH 256
QY 92 PQVNHSQLNESHKEDLOREHKSQSKSDTKVTATVLD-----KNNISKS 137
Db 257 KDINS-----DNQNKDINSDDNH--NDSTNESITTSLSINNTNRNKKRKNINNN 311
QY 138 TTNNPN 143
Db 312 NNNNSN 317
RESULT 26
ID ABP40312 standard; Protein; 902 AA.
XX
AC ABP40312;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5157.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
FN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
XX 08-NOV-1997; 97US-064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX WPI; 2002-381255/41.
XX N-PSDB; ABN92857.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS Disclosure; SEQ ID 5157; 267pp; English.
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
XX frame (ORF) nucleic acid sequences which encode the amino acid sequences
XX given in ABP35124 to ABP37960. The S. epidermidis sequences have
XX antibacterial activity and can be used in gene therapy. The sequences
XX can also be used in the diagnosis and treatment of bacterial infections,
XX particularly S. epidermidis infections. The sequences can be used to
XX screen for compounds able to interfere with the S. epidermidis life
XX cycle or inhibit S. epidermidis infection.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX USPTO web site.
XX
SQ Sequence 902 AA;
Query Match 11.6%; Score 87; DB 23; Length 902;
Best Local Similarity 25.3%; Pred. No. 7.8;
Matches 38; Conservative 27; Mismatches 65; Indels 20; Gaps 7;
```





PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 16-AUG-1999; 99US-0148684.  
PR 17-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 04-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.



```
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160988.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.5%; Score 86; DB 21; Length 408;
Best Local Similarity 23.6%; Pred. NO. 3.3;
Matches 35; Conservative 28; Mismatches 55; Indels 30; Gaps 7;

Qy 1 HRVTVIQNGEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGY--VINL 58
Db 209 HPVSALFYANERRAALREKNKSVVEVAK-----ITGEWNKLSDKKAPYKVAKK 260
Qy 59 SKDTFIKPVFKIEKKEENKPTFDVSKKDNPOVNHSQLNESHKREDLQREHSOKSD 118
Db 261 NKEYILO-AMEYKRTKEE-----ALSQKE-----EEELKLHKQALQWLKKEKTD 309
Qy 119 SKDQVATVLDKNNISSTNN--PNK 144
Db 310 N-----LIKKEKATKKKKNNVDPNK 330

RESULT 31
AAG37133
ID AAG37133 standard; Protein; 456 AA.
XX
AC AAG37133;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45609.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135134.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
```



PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0137282.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 07-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.





PA (SINO-) SINGENOMAX CO LTD.  
 XX Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;  
 XX WPI: 2001-355877/37.  
 DR N-PSDB; AAH62803.  
 XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus  
 PT (WSBV), useful for producing viral polypeptides that can be used to  
 PT screen for agents that are useful for treating WSBV infection -  
 XX Claim 1; Figure 3; 626pp; English.  
 XX The invention provides the primary nucleotide sequence of the WSBV genome  
 CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and  
 CC encoded proteins (AAH62840-63160) suitable for use as primers or probes. The nucleic acid  
 CC molecules and proteins of the invention are useful for diagnosis and  
 CC monitoring viral infection, in screens for antiviral agents and for  
 CC monitoring viral gene expression or activity during a treatment regimen.  
 CC The nucleic acid molecules are also useful as antisense constructs to  
 CC control viral gene expression in infected cells and tissues and to create  
 CC transgenic viral resistant shrimp.  
 XX Sequence 922 AA;  
 SQ  
 Query Match 11.5%; Score 86; DB 22; Length 922;  
 Best Local Similarity 21.9%; Pred. No. 10;  
 Matches 39; Conservative 28; Mismatches 63; Indels 48; Gaps 7;  
 QY 3 VVTWQNG-----KEMSTIVSEDFILPVY-----KGLR 34  
 DB 310 VTESMENGCCCFKNDKWLAKRESNLNNTVFGEDDEKSAAYVSDSEDEDENEEV 369  
 QY 35 GYQFDCWEISGEGKKDAGVNLKSDTIKPV-FKIEEKKKEENKPTFDVSKKKNPQ 93  
 DB 370 DDYNNETIESSVG-----NVKNLRKIGLSDVEEKEEGEQSEEEEDSDDDDD 421  
 QY 94 VNHSQNLNESHKEDL-----QREHSQKSDSTKQVATVL-----DKNNISSKSTNN 141  
 DB 422 DASSVCSSSSSSVTVAAAEEDDEEDKDKD-KATVWEDEDDKESVISSSSDS 478  
 RESULT 36  
 AAY08642  
 ID AAY08642 standard; Protein; 1315 AA.  
 XX AAY08642;  
 AC AAY08642;  
 XX AAY08642;  
 DT 20-MAR-2003 (updated)  
 DT 09-AUG-1999 (first entry)  
 XX S. aureus SdrD protein.  
 DE S. aureus SdrD protein.  
 XX Fibrinogen-binding protein; alpha chain; beta chain; ClfB; SdrC; SdrD;  
 KW SdrE; fibrinogen; medical device; competitive inhibitor; pharmaceutical;  
 KW treatment; infection; septicemia; osteomyelitis; mastitis; endocarditis;  
 KW extracellular matrix; vascular graft; vascular stent; vaccine;  
 KW intravenous catheter; artificial heart valve; cardiac assist device;  
 KW antibacterial.  
 XX Staphylococcus aureus.  
 OS Staphylococcus aureus.  
 XX WO9927109-A2.  
 PN WO9927109-A2.  
 XX 03-JUN-1999.  
 PD 03-JUN-1999.  
 XX 25-NOV-1998; 98WO-0525246.  
 PF 25-NOV-1998; 98WO-0525246.  
 XX 31-AUG-1998; 98US-0098427.  
 PR 31-AUG-1998; 98US-0098427.  
 XX 26-NOV-1997; 97US-0066815.  
 DE 26-NOV-1997; 97US-0066815.  
 XX (INHIT-) INHIBITEX INC.  
 PA

PA (FORF-) FORFAS T/A BIORESEARCH IRELAND.  
 PA (TEXA) UNIV TEXAS A & M.  
 PA (PATT/) PATTI J M.  
 PA (FOS/) FOSTER T J.  
 PA (JOSE/) JOSEFSSON E.  
 PA (EIDH/) EIDHIN D N.  
 PA (HOOK/) HOOK M A O.  
 PA (PERK/) PERKINS S E.  
 XX Bidhin DN, Foster TV, Hook MAO, Josefsson E, Patti JM;  
 PI Perkins SE;  
 XX WPI: 1999-357844/30.  
 DR N-PSDB; AAX77593.  
 XX Staphylococcus aureus fibrinogen-binding proteins for treating  
 PT septicemia, osteomyelitis, mastitis or endocarditis  
 PT Claim 8; Fig 8; 143pp; English.  
 PS This invention describes novel Staphylococcus aureus fibrinogen-binding  
 CC proteins that bind both the alpha and beta fibrinogen chains. The  
 CC proteins (and their encoding nucleic acids are ClfB, SdrC, SdrD and  
 CC SdrE). Staphylococcus aureus is thought to utilize fibrinogen to adhere  
 CC to medical devices, binding proteins that bind both the alpha and beta  
 CC fibrinogen chains (ClfB, SdrC, SdrD and SdrE) can therefore be used as  
 CC competitive inhibitors to block this binding. Antibodies against ClfB,  
 CC SdrC, SdrD and SdrE inhibit ClfB, SdrC, SdrD and SdrE mediated binding.  
 CC The proteins of the invention can be used in a pharmaceutical composition  
 CC for the treatment of Staphylococcus aureus infection e.g. septicemia,  
 CC osteomyelitis, mastitis or endocarditis or to inhibit the binding of  
 CC S. aureus to the extracellular matrix. The proteins or their fragments  
 CC may be used to coat a medical device to reduce the S. aureus infection of  
 CC an indwelling medical device, especially where the medical device is  
 CC selected from the group consisting of vascular grafts, vascular stents,  
 CC intravenous catheters, artificial heart valves, and cardiac assist  
 CC devices. ClfB, SdrC, SdrD, SdrE, or an active fragment, subdomain or  
 CC encoding gene may be used as a vaccine. The DS (aspartate serine) repeat  
 CC region or a gene encoding it may be used as an identifying probe for the  
 CC identification of genes and encoding proteins from Staphylococcus aureus  
 CC (other than ClfA), S. hemolyticus, S. lugdenensis, and S. schleriferi.  
 CC The proteins of the invention have antibacterial activity.  
 CC (Updated on 20-MAR-2003 to correct PA field.)  
 XX Sequence 1315 AA;  
 SQ  
 Query Match 11.5%; Score 86; DB 20; Length 1315;  
 Best Local Similarity 24.2%; Pred. No. 17;  
 Matches 36; Conservative 22; Mismatches 65; Indels 26; Gaps 4;  
 QY 3 VVTWQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGP--EGKDKAGYVNLK 60  
 DB 819 VVTWLKN-----ENGEVLQTTKDKGKYQFTGLENGTYKVEFETPSGYTPT--- 865  
 QY 61 DTFIKPVFKIEEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDST 120  
 DB 866 -----QVSGTGDEGIDNSGTSTGVKDKNDTDSGTGFKPTYNLGDYVWEDINKNGVQD 920  
 QY 121 KD-----VTATVLDKNNISSKSTNNPN 143  
 DB 921 KDEKGISGVTVTLKDENDKVLKTVTTDEN 949  
 RESULT 37  
 ABJ18969  
 ID ABJ18969 standard; Protein; 1315 AA.  
 XX ABJ18969;  
 AC ABJ18969;  
 XX 06-MAR-2003 (first entry)  
 DT 06-MAR-2003 (first entry)  
 XX Pathogen specific antigen related staphylococcal protein SEQ ID No 124.  
 DE Pathogen specific antigen related staphylococcal protein SEQ ID No 124.  
 XX



```

QY 121 KD-----VTATVLDKNNISSKSTNNPN 143
DB 921 KDEKGISGVTTLKDENDKVLKVTVDEN 949

RESULT 39
AAU37544
ID AAU37544 standard; Protein; 1349 AA.
AC AAU37544;
DT 14-FEB-2002 (first entry)
DE Staphylococcus aureus cellular proliferation protein #1714.
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
OS Staphylococcus aureus.
XX WO200170955-A2.
XX 27-SEP-2001.
XX 21-MAR-2001; 2001WO-US09180.
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 12-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX N-PSDB; AAS55403.
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX Example 3; Seq ID No 13137; 51pp; English.
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence represents an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 1349 AA;
XX Query Match 11.5%; Score 86; DB 22; Length 1349;
XX Best Local Similarity 24.2%; Pred. No. 17;
XX Matches 36; Conservative 22; Mismatches 65; Indels 26; Gaps 4;

QY 3 VTTVIONGKMSSTIVSEEDFILPVYKGELEKGYQDCWEISGF--EGKXDAGYVINLSK 60
DB 819 VTTVLKN-----ENGEVLQTTKTDKQKGYQFTGLENGTYKVEFETSGYTPT--- 865
QY 61 DTFIKPVFKKIEEKEENKPTFDVSKKKDPQVNHSQLNESHKREDLQREHHSQKSDST 120
DB 866 -----QVSGTDEGIDSNSTTGVIKDKNDTIDSGFYKPTYNLGDYVWEDTNKNGVQD 920
QY 121 KD-----VTATVLDKNNISSKSTNNPN 143
DB 921 KDEKGISGVTTLKDENDKVLKVTVDEN 949

RESULT 40
AAB23563
ID AAB23563 standard; Protein; 1980 AA.
AC AAB23563;
XX 09-JAN-2001 (first entry)
DE Human sodium channel 8A (SCN8A) protein sequence.
KW Sodium channel 8A; SCN8A; human; excitatory cell.
XX Homo sapiens.
XX JP2000201684-A.
XX 25-JUL-2000.
XX 11-JAN-1999; 99JP-0004645.
XX 11-JAN-1999; 99JP-0004645.
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX WPI; 2000-545978/50.
XX N-PSDB; AAA93791.
XX New sodium channel SCN8A useful for the elucidation physiological
XX mechanism participated by excitatory cells -
XX Disclosure; Page 5-12; 14pp; Japanese.
XX This sequence represents the human sodium channel SCN8A protein. The
XX invention also provides the SCN8A encoding gene sequence and an antibody
XX against the sodium channel. The SCN8A protein alpha subunit and its
XX encoding cDNA can be used in the elucidation of the physiological
XX mechanisms of excitatory cells.
XX Sequence 1980 AA;
XX Query Match 11.5%; Score 86; DB 21; Length 1980;
XX Best Local Similarity 26.2%; Pred. No. 29;
XX Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;
QY 37 QPDGWEISGFEGKDGAGYVINLSKDTFIKVPFKIEEKEENKPTFDVSKKKDPQVNH 96
DB 993 ENNNLIQISVIRIKGVAMT-KLVHAFMQAHFK---QREADEVKPLDELYEKKANCIAH 1048
QY 97 SOLNESHKREDLQREHHSQKSDSTKDVATVLDKNNISSKSTNNPN 143
DB 1049 TGA-DIHRNGDFQKNGTTSIGSGSVEKYIIDEDHM---SFINNPN 1091

Search completed: February 10, 2004, 10:53:53
Job time : 33.9025 secs

```



Result No.	Query Match	Score	Length	DB	ID	Description
1	97.5	13.0	1345	1	YH00 YEAST	P38800 saccharomyc
2	95	12.7	348	1	CYL2 HUMAN	Q14093 homo sapien
3	95	12.7	1398	1	TOP2 PLAFK	P41001 plasmodium
4	92.5	12.4	2468	1	MAPB HUMAN	P46821 homo sapien
5	90.5	12.1	1332	1	SPY7 YEAST	P35177 saccharomyc
6	89	11.9	1888	1	YD72 SCHPO	O14207 schizosacch
7	87.5	11.7	1702	1	IGA2 HAEIN	P45384 haemophilus
8	87	11.6	893	1	GYRA STAEF	P54112 staphylococ
9	87	11.6	2464	1	MAPB MOUSE	P14873 mus musculu
10	86.5	11.6	1006	1	RAT1 YEAST	Q02792 saccharomyc
11	86	11.5	443	1	GLNA PYRKO	O08467 pyrococcus
12	85.5	11.4	433	1	TIG STAAH	Q991i6 staphylococ
13	85.5	11.4	1202	1	PD0M ASCIM	P22374 ascobolus i
14	85	11.4	720	1	IF2 STAEF	Q8c6t4 staphylococ
15	84.5	11.3	1694	1	IGA0 HAEIN	P44969 haemophilus
16	84	11.2	778	1	YF8 HAEIN	P43610 saccharomyc
17	84	11.2	5596	1	MDN1 HUMAN	Q9nu22 homo sapien
18	83.5	11.2	479	1	U2r1 HUMAN	O15695 homo sapien
19	83	11.1	258	1	IF31 HUMAN	O75822 homo sapien
20	83	11.1	270	1	TONB HAEIN	P42872 haemophilus
21	83	11.1	688	1	LIP STAEF	Q02510 staphylococ
22	83	11.1	715	1	ADSV HUMAN	Q9Y6u3 homo sapien
23	82.5	11.0	439	1	GLNA PYRAB	Q9uy99 pyrococcus
24	82.5	11.0	847	1	DNLI CAEEL	Q27474 caenorhabdi
25	82.5	11.0	943	1	ARB2 DROME	Q79vk7 drosophila
26	82.5	11.0	2459	1	MAPB RAT	P15205 rattus norv
27	82	11.0	279	1	YMB4 YEAST	P49957 saccharomyc
28	82	11.0	427	1	YKTS YEAST	P36046 saccharomyc
29	82	11.0	439	1	GLNA PYRFU	Q05907 pyrococcus
30	82	11.0	439	1	GLNA PYRWO	P36687 pyrococcus
31	82	11.0	914	1	BPBA BACSU	P37993 bacillus su
32	82	11.0	949	1	IF2 HELPJ	Q9zm46 helicobacte
33	81.5	10.9	572	1	LM01 HUMAN	P29536 homo sapien

```
Db 1150 SHDKRPPFSKVE-----QKSSSRKSDNDKDLTHLDFVQNNFSEIFMKNLLSP 1201
QY 143 NK 144
Db 1202 QK 1203

RESULT 2
CYL2_HUMAN
ID _CYL2_HUMAN STANDARD; PRT; 348 AA.
AC Q14053;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cyclicin II (Multiple-band polypeptide II).
GN CYLC2 OR CYL2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=95255491; PubMed=7737358;
RA Hess H., Heid H., Zimbelmann R., Franke W.W.;
RT "The protein complexity of the cytoskeleton of bovine and human sperm
RT heads: the identification and characterization of cyclicin II.";
RL Exp. Cell Res. 218:174-182(1995).
CC -!- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
CC BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC -!- TISSUE SPECIFICITY: Testis.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46788; CAA86752.1; -
DR PIR; I37271; I37271.
DR Genew; HGNC:2583; CYLC2.
DR MIM; 604035; -
DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT DOMAIN 25 347 31 X 3 AA REPEATS OF K-K-X.
FT REPEAT 157 240 3 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 157 184 1.
FT REPEAT 185 212 2.
FT REPEAT 213 240 3.
SQ SEQUENCE 348 AA; 39079 MW; D86765599C1809E7 CRC64;

Query Match 12.7%; Score 95; DB 1; Length 348;
Best Local Similarity 30.9%; Pred. No. 2.1;
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 29 KGELEKGYQFDGWEISGFEGKDGAGVNLKDTFKIPVKFKIEKKEENKPTFP---DV 85
Db 205 ESEGEKG-----GTEKDSKKGKDS-----KKGKDSAIELQAVKDEKDDGKKDKGDE 256

QY 86 SK--KKDNPQVNHSQLN-----ESHKREDLQREHSHOKSDSTKD---VTATVLDKNNI 133
Db 257 SKDAKKDAKEIKGKKDKKKPSTSDSKDDVAKK-----SKDATKDAKKVAKKOTEKESA 313

QY 134 SSK 136
Db 314 DSK 316

RESULT 3
TOP2_PLAFK
ID TOP2_PLAFK STANDARD; PRT; 1398 AA.
AC P41001;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94316496; PubMed=8041616;
RA Chessman S., McAleese S., Goman M., Johnson D., Horrocks P.,
RA Ridley R.G., Kilbey B.J.;
RT "The gene encoding topoisomerase II from Plasmodium falciparum.";
RL Nucleic Acids Res. 22:2547-2551(1994).
CC -!- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -!- SIMILARITY: Belongs to the type II topoisomerase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X79345; -; NOT_ANNOTATED_CDS.
DR HSSP; P06786; 1BGW.
DR InterPro; IPR003594; ATPbind ATPase.
DR InterPro; IPR003957; CBFA_NFVB_topis.
DR InterPro; IPR001241; DNA_topoisolv.
DR InterPro; IPR002205; DNA_topoisolv.
DR Pfam; PF00204; DNA_gyraseB; 1.
DR Pfam; PF00521; DNA_topoisolv; 1.
DR Pfam; PF02518; HATPase_c; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PRINTS; PR00418; TP12FAMILY.
DR PRODOM; PD000742; DNA_topoisolv; 1.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00433; TOP2c; 1.
DR SMART; SM00434; TOP4c; 1.
DR PROSITE; PS00177; TOPOISOMERASE II; 1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
FT NP_BIND 144 149 ATP (POTENTIAL).
FT ACT_SITE 830 830 DNA CLEAVAGE (BY SIMILARITY).
FT DOMAIN 271 281 POLY-ASN.
FT DOMAIN 308 316 POLY-ASN.
FT DOMAIN 1089 1093 POLY-LYS.
FT DOMAIN 1227 1234 POLY-LYS.
SQ SEQUENCE 1398 AA; -161029 MW; BAAD7BEE80FE5BE9 CRC64;

Query Match 12.7%; Score 95; DB 1; Length 1398;
Best Local Similarity 23.8%; Pred. No. 9.4;
Matches 36; Conservative 36; Mismatches 36; Indels 26; Gaps 6;

QY 9 NGKEMSTTVSEEDFIL--PVYKGELEKGYQFDGWEISGFEGKDGAGVNLKDTFKIP 66
Db 1133 NEEIAGITVKDYDYLKSNPISLTLEK---VEDLLTLQKEKELEILRNITVTWMLK 1189

QY 67 VPKKIEE-----KKEENKPTFDVSKKKNPQVNHSQLNESHKREDLQREHSHQ 115
```

Db 1190 DIEKVEAIEPQNVNELNRESNK--PKVARKQ-----GPSMKKKKKKKLSSDSESE 1242

Qy 116 ---KSDTKDVTATVLDKNNISSKSTNNPN 143

Db 1243 GGDTSDSSEFLVTLNLIKNTTKTTSSNN 1273

RESULT 4

MAPB HUMAN

AC P46821; STANDARD; PRT; 2468 AA.

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain LC1].

DE LC1].

GN Homo sapiens (Human).

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OC NCBI\_TaxID=9606;

OX (1)

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal brain;

RX MEDLINE=95104835; PubMed=7806212;

RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;

RT "Cloning of human microtubule-associated protein 1B and the identification of a related gene on chromosome 15.";

RL Genomics 22:273-280(1994).

CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN. PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN STABILIZING MICROTUBULES.

CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE WITH MAP1A AND MAP1B PROTEINS.

CC -!- DOMAIN: Has a highly basic region with many copies of the sequence KEE and KKEI/V repeated but not at fixed intervals, which is responsible for the binding of MAP1B to microtubules.

CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated from MAP1B by proteolytic processing. It is free to associate with both MAP1A and MAP1B. It interacts with the amino-terminal region of MAP1B (By similarity).

CC -!- SIMILARITY: TO MAP1A.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; L06237; AAA18904.1; .

DR Genew; HGNC:6836; MAP1B.

DR MIM; 157129; .

DR GO; GO:0005875; C:Microtubule associated complex; TAS.

DR InterPro; IPR000102; MAP1B neuraxin.

DR Pfam; PF00414; MAP1B neuraxin; 10.

DR PROSITE; PS00230; MAP1B NEURAXIN; 6.

KW Microtubules; Repeat; Phosphorylation.

FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.

FT REPEAT 1878 1894

FT REPEAT 1895 1911

FT REPEAT 1912 1928

FT REPEAT 1929 1945

FT REPEAT 1946 1962

FT REPEAT 1963 1979

FT REPEAT 1997 2013

FT REPEAT 2014 2030

FT REPEAT 2031 2047

FT REPEAT 2048 2064

FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY KKEE AND KKEI/V REPEATS).

ST SEQUENCE 2468 AA; 270618 MW; 540839CDBF09D461 CRC64;

Query Match 12.4%; Score 92.5; DB 1; Length 2468;

Best Local Similarity 31.0%; Pred. No. 26;

Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

Qy 58 LSKDTFIKVPFKLEKEEENKPTFDVSKKNDPQVNHSQLNESHKEDLQRE-----E 112

Db 638 VKKETVKVP-----EDKKEEKEKPKKEVAKEDKTPI---KKEEKPKEEYKKEVKKEIK 689

Qy 113 HSQKSDSTKDV 123

Db 690 KEEKKEPKKEV 700

RESULT 5

SPT7 YEAST

ID SPT7 YEAST STANDARD; PRT; 1332 AA.

AC P35177;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Transcriptional activator SPT7.

GN SPT7 OR YBR081C OR YBR0739.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=95225044; PubMed=7713415;

RA Gansheroff L.J., Dollard C., Tan P., Winston F.;

RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein important for transcription in vivo.";

RL Genetics 139:523-536(1995).

CC [2]

CC SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=95076715; PubMed=7985423;

RA van der Aart Q.J.M., Barthe C., Doignon F., Aigle M., Crouzet M., Steensma H.Y.;

RT "Sequence analysis of a 31 kb DNA fragment from the right arm of Saccharomyces cerevisiae chromosome II.";

RL Yeast 10:959-964(1994).

CC [3]

CC SEQUENCE OF 1-835 FROM N.A.

RC STRAIN=S288C;

RX Vissers S., Cziepluch C., Hein C., Jauniaux J.C., Urrestazu A., Andre B., Vissers S.;

RT Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.

CC [4]

CC SEQUENCE OF 463-523 FROM N.A.

RX MEDLINE=92285152; PubMed=1350857;

RA Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J., David I.B.;

RT "The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins.";

RL Nucleic Acids Res. 20:2603-2603(1992).

CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY OTHER GENES.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- SIMILARITY: Contains 1 bromodomain.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

```

CC -----
DR EMBL; L22537; AAC37424.1; -
DR DR EMBL; X76294; CAA53940.1; -
DR DR EMBL; X35950; CAA85026.1; -
DR DR EMBL; M87651; AAA35087.1; -
DR DR PIR; S41552; S41552.
DR DR HSP; Q92831; I891.
DR DR TRANSFAC; T04835; -
DR DR SGD; S0000285; SPT7.
DR DR GO; GO:000124; C:SAGA complex; IDA.
DR DR InterPro; IPR001487; Bromodomain.
DR DR Pfam; PF00439; bromodomain; 1.
DR DR PRINTS; PR00503; BROMODOMAIN.
DR DR SMART; SM00297; BROMO; 1.
DR DR PROSITE; PS00633; BROMODOMAIN 1; 1.
DR DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR KW Transcription regulation; Nuclear protein; Activator; Bromodomain.
FT DOMAIN 458 528 BROMODOMAIN.
SQ SEQUENCE 1332 AA; 152616 MW; 083B63624669244F CRC64;

Query Match 12.1%; Score 90.5; DB 1; Length 1332;
Best Local Similarity 24.5%; Pred. No. 19;
Matches 39; Conservative 26; Mismatches 57; Indels 37; Gaps 8;

Qy 5 VTQNGKMSSTIVSEEDFILPVYKGEIKGYQPDGWEISGFEGKQDAGYVNLNSKDTFI 64
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 549 ITINRADLEKEI---EDM-----EKDKDYELDEEEVAGSRKG-----LNMGAHMLA 594
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :

Qy 65 K---PVPFKIEEKKKEENKPTFD-----VSKKXD-----NPQVNHSQLNESH 103
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 595 KENGKGVSEKDSKTKVQDEAPTNDDKLTSVIPEGEKEKDKTASSTVTVTNENVKNKIEKENG 654
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :

Qy 104 RKEDLQR-EHHSQKSDSTKQVATVLD-KNINISKSTTN 140
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :
Db 655 KNEQDNVESSSKTSDSSKDAAKKDTDEGLQDKTAEN 693
   ||| : : ||| : : ||| : : ||| : : ||| : : ||| : :

```

QY	14	SSITVSEDF-----ILPVY---KGELEKGYQDFDGEWISGFE-----GKKDAG 53
Db	390	SSLSITSENPFOLNVAANAVSTIPVYRTTKTKMKN-RFKYVEVEKULPDILESYGKAPK 448
QY	54	YVINLSKDTIKPFVKFIEKEEENKPTPDVSKKKDNQVNHSSQLNESHKEDLQREH 113
Db	449	FLRVFARSSSHIP--KWIREKQWDSKKYFSPDKESDRQVIDQVLSWYSGKHQLVQOSH 206
QY	114	SQKSDS-TKDVVTATVLDKN-----NISKSTNN 141
Db	507	SYKKPSDSKSGVGNIFSVNSKHSVINAKTAANN 541
RESULT 7		
ID	IGA2	HAELIN STANDARD; PRT; 1702 AA.
AC	P45384;	
DT	01-NOV-1995	(Rel. 32, Created)
DT	01-NOV-1995	(Rel. 32, Last sequence update)
DT	28-FEB-2003	(Rel. 41, Last annotation update)
DE	Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).	
GN	IGA1	
OS	Haemophilus influenzae.	
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;	
OC	Pasteurellaceae; Haemophilus.	
OX	NCBI_TaxID=727;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=HK715 / Serotype B;	
RX	MEDLINE=92234949; PubMed=1373717;	
RA	Poulsen K., Reinholdt J., Kilian M.;	
RT	"A comparative genetic study of serologically distinct Haemophilus	
RT	influenzae type 1 immunoglobulin A1 proteases."	
RL	J. Bacteriol. 174:2913-2921(1992).	
CC	-!- PRODUCING INTACT FC AND FAB FRAGMENTS IMMUNOGLOBULIN A	





```

FT BINDING      362      362      AMP (UNDER CONDITIONS OF ABUNDANT
FT              443 AA; 50259 MW; 9426DCCFEF18168 CRC64;
SQ SEQUENCE      11.5%; Score 86; DB 1; Length 443;
Query Match      Best Local Similarity 17.8%; Pred. No. 13;
Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;

Qy 28 YKGELEKGYQFDGNETSGFPGKKDAGVYNLSKDTFI----- 64
   | : | : | | | | | : | : | : | : | : | : | : | : |
Db 42 YEEAVEDGVSGFDGSSIPGFGEIDSDLIKFADPSTVAEIPWEGIGRVGYIYKGDPEYQA 101
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 65 --KPVFKKIEEKKEEEN-----KPTFDVSKKKD----- 90
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 102 DPGILKRVLEULEKGLKKAHIGPEPFYIFKQGTWELHIPDGGYFDLVGLDKAREIR 161
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 91 -----NPQVNHSQLNESHKEDLQREHSQKSD---STKDVTATVLD 129
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 162 RETALYMPYIGLKPEVLIHHEVGKAAOHEIDPRYDEALRTADNIYVFHKHVKAAVE 215
   | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

TIG STAM

AC  
ITG

DT	28-87
DT	29-87

DT 28-87

GN TIG 8

OS  
Staple

OC Bact

RN [1]

RC STRA:

RA Kuroo

RA Kanar

RA Sekin

RA  
HartRT  
all over  
MIOMI

RT  
BN

RP  
D

RX MEDL.

RA Naga:

RT "Gend

RL Lance

22

500

cc This

the I

modi:

33 33

CC  
DD  
---  
EMPT



```
DR EMBL; AB003134; BAB42766.1; -.
DR EMBL; AF004827; BAB95484.1; -.
DR PIR; A89951; A89951.
DR HAMAP; MF 00303; -. 1.
DR InterPro; IPR001179; FKBP_PP1ase.
DR InterPro; IPR005215; Trig_fac.
DR Pfam; PF00254; FKBP; 1.
DR TIGRFAMs; TIGR00115; tlg; 1.
DR PROSITE; PS00453; FKBP_PP1ASE_1; FALSE NEG.
DR PROSITE; PS00454; FKBP_PP1ASE_2; FALSE NEG.
DR PROSITE; PS00059; FKBP_PP1ASE_3; 1.
DR Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.
KW DOMAIN 163 248 PP1ASE, FKBP-TYPE.
FT SEQUENCE 433 AA; 48609 MW; 8865D9AF6A1BC1E7 CRC64;

Query Match 11.4%; Score 85.5; DB 1; Length 433;
Best Local Similarity 22.9%; Pred. No. 14;
Matches 49; Conservative 30; Mismatches 52; Indels 83; Gaps 13;

QY 3 VTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 34
Db 93 VSVTQIEKGKDFTEATVPEVEVGLDYGKLEIEKQETELSDDELQEAIDHSLGLAEM 152
QY 35 -----GYQFGWGEISG--PEGKXGAGVYVNSKDTFKPVFK-----KIE 72
Db 153 VKKEDGVNGDTVNIIDFSG-SVDGEFEGGQAGYDLEIGSGSFT-PGFESQLEGKMYD 210
QY 73 EKKE-----EE--NKPTFVDS-----KKDNQPQVNSQLNE-----SHRKE 106
Db 211 EKKDVVTPPEVHAELAGKATFTKYNEIKFEVPELTDEIANELDAEANTVDVEYKE 270
QY 107 DLQREHSQKSDTKDVTATVLDKNNISSKSTTN 140
Db 271 NLRKRLAEQKATDAENV-----EKEAITKATDN 299

RESULT 13
DPOM ASCIM
ID DPOM ASCIM STANDARD; PRT; 1202 AA.
AC P22374;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Probable DNA polymerase (EC 2.7.7.7).
OS Ascobolus immersus.
OG Mitochondrion.
OG Plasmid PA12.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Pezizomycetes;
OC Pezizales; Ascombolaceae; Ascobolus.
OX NCBI_TaxID=5191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2/I;
RX MEDLINE=90066356; PubMed=2573821;
RA Kempken F., Meinhardt F., Esser K.;
RT "In organello replication and viral affinity of linear,
RT extrachromosomal DNA of the ascomycete Ascobolus immersus.";
RL Mol. Gen. Genet. 218:523-530(1989).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA] (N).
CC -1- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG, TO
CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB016747; AAO04542.1; -.
CC HAMAP; MF 00100; -. 1.
CC Pfam; PF00009; GTP_EFTU; 1.
CC Pfam; PF03144; GTP_EFTU_D2; 1.
CC Pfam; PF04760; IF2_N; 2.
CC Pfam; PF04760; IF2_N; 2.
CC ProDom; PD186100; IF2; 1.
CC TIGRFAMs; TIGR00487; IF-2; 1.
CC TIGRFAMs; TIGR00231; small_GTP; 1.
CC PROSITE; PS01176; IF2; 1.
CC Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.

DR EMBL; X15982; CAA34106.1; -.
DR PIR; S05362; S05362.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR004868; DNA_pol_B_2.
DR Pfam; PF03175; DNA_pol_B_2; 1.
DR SMART; SM00486; POLBc; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Plasmid; Mitochondrion.
SQ SEQUENCE 1202 AA; 138279 MW; 51D41FCEBDBF2CDE CRC64;

Query Match 11.4%; Score 85.5; DB 1; Length 1202;
Best Local Similarity 22.5%; Pred. No. 41;
Matches 36; Conservative 31; Mismatches 64; Indels 29; Gaps 6;

QY 4 TVTIQNGKEMSSTIVSEEDFILPVYKGLEK-----GYQFGWGEISGFEKGDAGYVIN 57
Db 344 TLAIFLEREDITHVMSYDE-----GDIDSKFPKGSLSFDKPLKTIETGKYANYTFP 396
QY 58 LSKOTFIKPVFKIE-----EKKEENK-PTFQVSKKDNQVNSQLNESHKEDLQR 110
Db 397 IKKDIWVKDINKKINFNGLDLPKTMDSLKPWNLKDKTSGEIRMTIKNNKNSQSYDI-- 454
QY 111 BEHSQKSDTKDVTATVLDKNNI-----SKSSTTNPN 143
Db 455 IGHMIINDGENVITFNRADVNSIIKIFTVTDMSGNTNDPN 494

RESULT 14
IF2_STAEP
ID IF2_STAEP STANDARD; PRT; 720 AA.
AC Q8CST4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-2.
OS INF6 OR SE0945.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB016747; AAO04542.1; -.
CC HAMAP; MF 00100; -. 1.
CC Pfam; PF00009; GTP_EFTU; 1.
CC Pfam; PF03144; GTP_EFTU_D2; 1.
CC Pfam; PF04760; IF2_N; 2.
CC ProDom; PD186100; IF2; 1.
CC TIGRFAMs; TIGR00487; IF-2; 1.
CC TIGRFAMs; TIGR00231; small_GTP; 1.
CC PROSITE; PS01176; IF2; 1.
CC Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
```



```
FT DOMAIN 225 373 G-DOMAIN.
FT NP_BIND 231 238 GTP (BY SIMILARITY).
FT NP_BIND 277 281 GTP (BY SIMILARITY).
FT NP_BIND 331 334 GTP (BY SIMILARITY).
SQ SEQUENCE 720 AA; 79343 MW; 07FBSA6A59CF970C CRC64;

Query Match 11.4%; Score 85; DB 1; Length 720;
Best Local Similarity 28.0%; Pred. No. 26;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

Qy 71 IEEKKEENKTFDVSKKD-----NPQVNHSHKEDLQREHSQKSDSTK-----D 122
Db 38 LEEQIKALDKKFKASQAKDKNTQNNHQSNNKQNSNDKQKQNSKPTKKKEQN 97
Qy 123 VTATVLDKNNISSKSTTNPNK 144
Db 98 NKGQKQNNKNTKNNKNNK 119

RESULT 15
ID_IGA0_HA0IN STANDARD; PRT; 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease).
GN IGA OR IGAL OR HI0990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=serotype D;
RA Wright A., Fishman Y., Tai F., Plaut A.G.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; X59800; -, NOT ANNOTATED CDS.
DR EMBL; U32779; AAC22651.1; -.
DR PIR; H64106; H64106.
DR MEROPS; S06.001; -.
DR TIGR; HI0990; -.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertactin.
DR Pfam; PF03797; Autotransporter; 1.
DR Pfam; PF02395; IGAL; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR TIGRFAMs; TIGR01414; autotrans_barl; 1.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal;
KW Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT CONFLICT 253 254 EN -> GV (IN REF. 1).
FT CONFLICT 272 272 G -> A (IN REF. 1).
FT CONFLICT 464 464 G -> E (IN REF. 1).
FT CONFLICT 866 866 S -> T (IN REF. 1).
FT CONFLICT 1036 1036 A -> D (IN REF. 1).
FT CONFLICT 1074 1074 A -> G (IN REF. 1).
FT CONFLICT 1421 1421 A -> G (IN REF. 1).
FT CONFLICT 1545 1545 H -> T (IN REF. 1).
SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;

Query Match 11.3%; Score 84.5; DB 1; Length 1694;
Best Local Similarity 26.1%; Pred. No. 70;
Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;

Qy 56 INLSKDTFIKPVFKIEEKKEENKPTFDVSKKDNQVNHSHKEDLQREHS- 114
Db 1288 INTGSATAITETAEKSKDPQTETAASTEDASQHKANTVADNSVANNSSSDPKSRRSI 1347
Qy 115 --QKSDSTKDVATVLDKNNISSKSTTNPNK 144
Db 1348 SQPQETSAEETTAASTDETTIADNSKRSKPNR 1379

RESULT 16
YFK8_YEAST
ID_YFK8_YEAST STANDARD; PRT; 778 AA.
AC P43610;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 88.7 kDa helicase in CDC26-SAP155 intergenic region.
GN YFR038W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae."
RL Nat. Genet. 10:261-268(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=96287654; PubMed=8686381;
RA Eki T., Naitou M., Hagiwara H., Abe M., Ozawa M., Sasanuma S.-I.,
RA Sasanuma M., Tsuchiya Y., Shibata T., Watanabe K., Ono A.,
```

RA Yamazaki M.-A., Tashiro H., Hanaoka F., Murakami Y.;  
 RT "Fifteen open reading frames in a 30.8 kb region of the right arm of  
 chromosome VI from *Saccharomyces cerevisiae*.";  
 RL Yeast 12:177-190(1996).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; D50617; BAA09277.1; -  
 DR PIR; S56293; S56293.  
 DR SGD; S0001934; YFR038W.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR000330; SNF2\_N.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00176; SNF2\_N; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELICG; 1.  
 KW Hypothetical protein; Nuclear protein; DNA-binding; Helicase;  
 KW ATP-binding.  
 FT NP\_BIND 247 254 ATP (POTENTIAL).  
 FT SITE 352 355 DSGH BOX.  
 SQ SEQUENCE 778 AA; 88730 MW; 3E6C0857B5EABD84 CRC64;  
 Query Match 11.2%; Score 84; DB 1; Length 778;  
 Best Local Similarity 25.5%; Pred. No. 33;  
 Matches 37; Conservative 18; Mismatches 56; Indels 34; Gaps 5;  
 QY 13 MSSTIVSEEDFILPVYKGEKGYQDFGWEISGFEGKDGAGVYVNLKDTFTKPVFKIE 72  
 DB 72 LQDVHSDEDI-----QLDSEDDSDTEAVQ-AQVVDKLAKDT-KSEQKSLD 115  
 QY 73 EKKEENKFTFVSKKKONPVNHLQNES-----HRKEDLQR-----EHS 114  
 DB 116 DELSEMDTKTVSLKLLKNEFRQSQVYSIIADTLLHRSNEVANANTKDNSDDEHS 175  
 QY 115 QKSDSTKDTATVLDKNISSKSTT 139  
 DB 176 SKKRKTKKKSITDFPKKQKNEDTT 200  
 RESULT 17  
 MDNI\_HUMAN  
 ID MDNI\_HUMAN STANDARD; PRT; 5596 AA.  
 AC Q9NU22; O15019;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Midasin (MIDAS-containing protein).  
 GN MDNI OR KIAA0301.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP TISSUE=Testis;  
 RX PubMed=12102729;  
 RA Garbarino J.E., Gibbons I.R.;  
 RT "Expression and genomic analysis of midasin, a novel and highly  
 conserved AAA protein distantly related to dynein.";  
 RL BMC Genomics 3:18-18(2002).  
 RN [2]  
 RP SEQUENCE OF 1255-2356 AND 3550-5596 FROM N.A.  
 RC TISSUE=Testis;  
 RA Tracey A.;

RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 3550-5596 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 The complete sequences of 100 new cDNA clones from brain which can  
 code for large proteins in vitro.";  
 RL DNA Res. 4:141-150(1997).  
 CC -1- FUNCTION: May function as a nuclear chaperone and be involved in  
 the assembly/disassembly of macromolecular complexes in the  
 nucleus.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: Contains 1 VWFA domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; AF503925; AAM77722.1; -  
 DR EMBL; AL096678; CAB86660.1; -  
 DR EMBL; AL096678; CAB86661.1; -  
 DR EMBL; AB002299; BAA20761.1; -  
 DR Genew; HGNC:18302; MDN1.  
 DR GO; GO:0005634; C:nucleus; NAS.  
 DR GO; GO:0016887; F:ATPase activity; NAS.  
 DR GO; GO:0003754; F:chaperone activity; NAS.  
 DR GO; GO:0006461; P:protein complex assembly; NAS.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00004; AAA; 1.  
 DR SMART; SM00382; AAA; 7.  
 DR SMART; SM00327; VWFA; 1.  
 DR PROSITE; PS50234; VWFA; 1.  
 KW Chaperone; ATP-binding; Repeat; Nuclear protein.  
 FT NP\_BIND 329 336 ATP (POTENTIAL).  
 FT NP\_BIND 677 684 ATP (POTENTIAL).  
 FT NP\_BIND 1094 1091 ATP (POTENTIAL).  
 FT NP\_BIND 1390 1397 ATP (POTENTIAL).  
 FT NP\_BIND 1753 1760 ATP (POTENTIAL).  
 FT NP\_BIND 2066 2073 ATP (POTENTIAL).  
 FT DOMAIN 3566 3573 POLY-GLU.  
 FT DOMAIN 4784 4791 POLY-GLU.  
 FT DOMAIN 5008 5013 POLY-GLU.  
 FT DOMAIN 5182 5187 VWFA.  
 FT DOMAIN 5384 5583 VWFA.  
 FT CONFLICT 2287 2312 RLFLSMDPVHGDISRMRNRLGIYI -> S (IN REF.  
 2).  
 SQ SEQUENCE 5596 AA; 632802 MW; 586C62616A1F96D4 CRC64;  
 Query Match 11.2%; Score 84; DB 1; Length 5596;  
 Best Local Similarity 23.2%; Pred. No. 2.7e+02;  
 Matches 29; Conservative 27; Mismatches 61; Indels 8; Gaps 2;  
 QY 1 HRVTVTIONGKMSSTIVSEEDFILPVYKGELEK-----YQFDGWEISGFEGKDGAGY 54  
 DB 4632 HRSTAKLSVLVAQVFTLAQKGFCLPKFEMDSAGEGATEFHVDYGGGIGEGEGMKDVS- 4690  
 QY 55 VINLSKDTFKIPVFKKIEKKEENKPTFDVSKKDNPNVNHSQLNESHKRDLOREHS 114  
 DB 4691 -DQIGNERQVEDTFQKGQEKEDPDSDIKGEDNAIEMSDPDGKVDHGLEBEQEDD 4749  
 QY 115 QKSDS 119  
 DB 4750 EKSDS 4754

```
RESULT 18
U2R1_HUMAN STANDARD; PRT; 479 AA.
AC Q15695; Q13570;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit
DE related-protein 1.
GN U2AF1-RS1 OR U2AF1RS1 OR U2AFBPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96163878; PubMed=8586425;
RA Kitagawa K., Wang X., Hatada I., Yamaoka T., Nojima H.,
RA Inazawa J., Abe T., Mitsuya K., Oshimura M., Murata A., Monden M.,
RA Mukai T.;
RT "Isolation and mapping of human homologues of an imprinted mouse gene
RT U2af1-rs1.";
RL Genomics 30:257-263(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96212931; PubMed=8630064;
RA Pearsall R.S., Shibata H., Brozowska A., Yoshino K., Okuda K.,
RA Dejong P.J., Platts C., Chapman V.M., Hayashizaki Y., Held W.A.;
RT "Absence of imprinting in U2AFBPL, a human homologue of the imprinted
RT mouse gene U2afbp-rs.";
RL Biochem. Biophys. Res. Commun. 222:171-177(1996).
CC -!- SURCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -!- SIMILARITY: Contains 2 C3H1-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D49676; BAA08532.1; -
CC EMBL; U51224; AAA98669.1; -
CC MIM; 601079; -
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0003723; F:RNA binding activity; NAS.
CC InterPro; IPR000504; RNA rec mot.
CC InterPro; IPR000571; Znf_CCH.
CC Pfam; PF00076; rrm; 1.
CC Pfam; PF00642; zf-CCH; 2.
CC SMART; SM00360; RRM; 1.
CC SMART; SM00356; Znf_C3H1; 2.
CC PROSITE; PS0102; RRM; 1.
CC PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW Nuclear protein; RNA-binding; Ribonucleoprotein; Zinc-finger; Repeat.
FT DOMAIN 203 309 RNA-BINDING (RRM).
SQ SEQUENCE 479 AA; 57643 MW; 96F326694BD4E7C0 CRC64;

Query Match 11.2%; Score 83.5; DB 1; Length 479;
Best Local Similarity 21.7%; Pred. No. 22;
Matches 26; Conservative 25; Mismatches 40; Indels 29; Gaps 4;

QY 50 KDAGVVLNLSKDTFKIPKVFKEKK-----EENKPTFDVSKKKNPQ 93
Db RSGLSQEEEDTFTIE--QLEEEKLLERERLHEEWLLREQQAQEFRIKKEEAA 101
QY 94 VN-----HSQNESHRKEDLQREHSQKSDSTKDVATVLD--KNNISKSTTNP 142
Db 102 KKWLEEQERKLKEQWKEQKEREEREQKEKEKEAVQKWLQDAENDLNSTTQNP 161
```

```
RESULT 19
IF31_HUMAN STANDARD; PRT; 258 AA.
AC O75822; Q98UD2; Q98Q2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Eukaryotic translation initiation factor 3 subunit 1 (eIF-3 alpha)
DE (eIF3 p35) (eIF3j).
GN EIF3S1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=99041954; PubMed=9822859;
RA Block K.L., Vornlocher H.-P., Hershey J.W.B.;
RT "Characterization of cDNAs encoding the p44 and p35 subunits of human
RT translation initiation factor eIF3.";
RL J. Biol. Chem. 273:31901-31908(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovarian carcinoma;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yanamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=22389257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buétow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Rouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA.
CC -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 12 DIFFERENT SUBUNITS (BY
CC similarity).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U97670; AAC78729.1; -
CC EMBL; AK023388; BAB14555.1; -
CC
```



RT "Molecular analysis and expression of the lipase of *Staphylococcus*  
 RL *epidermidis*.";  
 RN J. Gen. Microbiol. 139:267-277(1993).  
 RP [2]  
 RC SEQUENCE FROM N.A.  
 RA STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
 RA Chen Z., Wen Y.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
 CC fatty acid anion.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS  
 CC NEGATIVELY REGULATED BY BACTERIOPHAGE LYSOGENIZATION (LIPASE  
 CC CONVERSION).  
 CC -1- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M95577; AAA19729.1; --  
 CC EMBL; AE016744; AAO03878.1; --  
 CC PIR; A47705; A47705. Gpos\_Y5IRK.  
 CC InterPro; IPR005877; Gpos\_Y5IRK.  
 CC InterPro; IPR000734; Lipase.  
 CC InterPro; IPR000379; Ser esters site.  
 CC Pfam; PF04650; Y5IRK signal; 1.  
 CC TIGRFAMs; TIGR01168; Y5IRK signal; 1.  
 CC PROSITE; PS00120; LIPASE\_SER; 1.  
 CC HydroLase; Lipid degradation; Zymogen; Signal; Complete proteome.  
 KW PROPEP 36 302 REMOVED IN THE MATURE FORM.  
 FT SIGNAL 1 35 POTENTIAL.  
 FT CHAIN 303 688 LIPASE.  
 FT ACT\_SITE 418 418 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 648 648 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CONFLICT 96 96 W -> L (IN REF. 1).  
 FT CONFLICT 120 120 E -> G (IN REF. 1).  
 SQ SEQUENCE 688 AA; 77343 MW; 6C95DB3A78AF86F6 CRC64;  
 Query Match 11.1%; Score 83; DB 1; Length 688;  
 Best Local Similarity 24.5%; Pred. No. 35;  
 Matches 35; Conservative 28; Mismatches 60; Indels 20; Gaps 6;  
 Qy 6 TQNGKMSSTVSEEDFILPVYKGELEKGYQFDGWEISGFE-GKKDAGYVINLSKDTFI 64  
 Db 79 SITENSLHNETPKNEDWI-----QQQKDSQNDNKSSEVVEQNKENAFQVNHSE--- 129  
 Qy 65 KPVFKKIE-EKKKEENKPTFDVSKKDNQV--NHSQLNESHKEDLQREHSHQSDSTK 121  
 Db 130 KPQQQVELEKHAENNTLHSAQSNEDVKTSPQLDNTAAKQEDSKENLSKQDTQS 189  
 Qy 122 DVTATVLDKNISSSTNNPNK 144  
 Db 190 SKTTDLL-----RATAQNSK 205  
 RESULT 22  
 ADSV\_HUMAN  
 ID ADSV\_HUMAN STANDARD; PRT; 715 AA.  
 AC Q9Y6U3; Q8MU97;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Adseverin (Scinderin).  
 GN SCIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Pahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smillius D.E.,  
 RA Sutterch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE OF 1-527 FROM N.A.  
 RA Kalicki J., Smith-Craig R.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 248-715 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Isogai T., Oka T., Hayashi K., Sugiyama T., Otauki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Ca(2+)-dependent actin filament-severing protein that is  
 CC presumed to have a regulatory function in exocytosis by affecting  
 CC the organization of the microfilament network underneath the  
 CC plasma membrane. In vitro, also has barbed end capping and  
 CC nucleating activities in the presence of Ca(2+).  
 CC -1- SIMILARITY: BELONGS TO THE VILIN/GELSOLIN FAMILY.  
 CC -1- SIMILARITY: Contains 6 gelsolin-like repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; BC021090; AAU21090.1; --  
 CC EMBL; AC005281; AAD15423.1; --  
 CC EMBL; AK027778; BAB55361.1; --  
 CC HSSP; P02640; 2VIL.  
 CC InterPro; IPR001974; Gelsolin.  
 CC Pfam; PF00626; Gelsolin; 6.  
 CC PRINTS; PR00597; GELSOLIN.  
 CC SMART; SM00262; GEL; 6.  
 CC Cytoskeleton; Actin-binding; Repeat; Calcium; Capping protein.  
 KW DOMAIN 1 363 ACTIN-SEVERING (POTENTIAL).  
 FT DOMAIN 364 715 CA(2+)-DEPENDENT ACTIN BINDING.  
 FT REPEAT 27 76 GELSOLIN-LIKE 1.  
 FT REPEAT 148 188 GELSOLIN-LIKE 2.  
 FT REPEAT 265 307 GELSOLIN-LIKE 3.  
 FT REPEAT 398 451 GELSOLIN-LIKE 4.  
 FT REPEAT 523 564 GELSOLIN-LIKE 5.  
 FT REPEAT 626 668 GELSOLIN-LIKE 6.  
 FT SITE 112 119 POLYPHOSPHOINOSITIDE BINDING (BY

FT	SITE	138	146	SIMILARITY).
FT				POLYPHOSPHONOSITIDE BINDING (BY
FT				SIMILARITY).
FT	CONFLICT	61	61	R -> H (IN REF. 2).
SEQ	SEQUENCE	715 AA;	80508 MW;	45FBE42CCEBDD80 CRC64;
	Query Match	11.1%;	Score 83;	DB 1; Length 715;
	Best Local Similarity	23.1%;	Pred. No. 36;	
	Matches	34;	Conservative	30; Mismatches 47; Indels 36; Gaps 5;
QY		2	RVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGEGK-----KDAGYV 55	
DB		262	RVTVVAEENPFMSAMLLISECFILD--HGAAKQIFVWKGDKANPQERKAAMKTAEFLQQ 319	
QY		56	INLSKDTPIK-----PVFKK-----IEBKKEENKPTDFVSKKK 89	
DB		320	MNYSKNTQIQVLPGGGETPIFKQFKDWRDKQSDGFGKVYVTEKVAIQIKPIPFASKLH 379	
QY		90	DNPO--NHSQLNESHKEDLQREHS 114	
DB		380	SSPQMAAHNMVDDGSGKVEIWRVENN 406	
RESULT 23				
GLNA PYRAB				
ID	GLNA PYRAB	STANDARD;	PRT;	439 AA.
AC	Q9UY99;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).			
GN	OR PYRAB16090 OR PAB1292.			
OS	Pyrococcus abyssi.			
OC	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;			
OC	Pyrococcus.			
OX	NCBI_TaxID=29292;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GES / Orsay;			
RX	PubMed=12622808;			
RA	Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,			
RA	Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,			
RA	Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;			
RT	"An integrated analysis of the genome of the hyperthermophilic			
RT	archaeon Pyrococcus abyssi."			
RL	Mol. Microbiol. 47:1495-1512(2003).			
CC	-1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +			
CC	L-glutamine.			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).			
CC	-1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement">http://www.isb-sib.ch/announcement</a> ).			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .			
CC				
DR	EMBL; AJ248288; CAB50513.1; -			
DR	PIR; C75009; C75009.			
DR	HSP; P06201; ILGR.			
DR	InterPro; IPR001691; GLN synth.			
DR	InterPro; IPR004809; GlnA.			
DR	InterPro; IPR001637; GlnA_adenyltn.			
DR	Pfam; PF00120; gln-synt; 1.			
DR	Pfam; PF03951; gln-synt N; 1.			
DR	ProDom; PD001057; Gln synt C; 1.			
DR	TIGRFAMs; TIGR00653; GlnA_1.			
DR	PROSITE; PS00180; GLNA_1; 1.			
DR	PROSITE; PS00181; GLNA ATP; 1.			
DR	Ligase; Complete proteome.			
FT	BINDING	358	358	AMP (UNDER CONDITIONS OF ABUNDANT



```

DB 671 IRDDKNSDDATSSQ--VLEMYKNOFAFANQKIEKADAVD--EDDEFKEDDEEBELNMT 726
QY 59 -----SKDTFKPKVFKKIEEKEENKPTDFVSKK--DNPQVNHSLNESHKREDLQR 110
DB 727 NVSESSKENPVK-----EIKKSTPK--SVSPKFKPKPPVXSSPVNKSPPVKSSPIK 777
QY 111 EHSQKSDSTKDVATVLVDKNNISSKSTNNP 142
DB 778 KEAEKKGPVASIFSSTKQNEKQKVKVSPSP 809

RESULT 25
ARS2 DROME STANDARD; PRT; 943 AA.
AC Q9V9K7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Arsenite-resistance protein 2 homolog.
GN CG7843.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Aehburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazej R.G., Champe W., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
RA Foubin C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.M., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL SEQUENCE 287:2185-2195(2000).
RN [2]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RC STRAIN=Berkley;
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

```

```

RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Betencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q9V9K7-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q9V9K7-2; Sequence=VSP 000327;
CC -!- SIMILARITY: BELONGS TO THE ARS2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE003784; AAM68343.1; -
DR EMBL; AE003784; AAM68345.1; -
DR FlyBase; FBgn0033062; CG7843.
DR Pfam; PF04959; ARS2; 1.
KW Hypothetical protein; Alternative splicing.
FT VARSPLIC 47 50 Missing (in isoform Short).
FT FTID=VSP 000327.
SQ SEQUENCE 943 AA; 107221 MW; 0C1AF05E02E8AB0B CRC64;
Query Match 11.0%; Score 82.5; DB 1; Length 943;
Best Local Similarity 25.6%; Pred. No. 53;
Matches 32; Conservative 21; Mismatches 61; Indels 11; Gaps 3;
QY 7 IQNGKMSSTI-----VSEDFILPVVKGLEKGYQFDGWEISGFEGKDGAVYVNLK 60
DB 291 MQSVKEVKTINSKPEEMSEADPVSTOKPVPVNSDGENWDDDAENSAPKLAEDSK 350
QY 61 DTFIKPVKITEE---KKEENKPTDFVSKKKQNPQVNHSLNESHKREDLQREHSQKS 117
DB 351 DSDSKPEDKQLNKKTKRKKNSSDDSSSESSSSDEEKLKEKYDVEDGLRAE--QKT 408
QY 118 DSTKD 122
DB 409 EAERD 413
RESULT 26
MAPB RAT STANDARD; PRT; 2459 AA.
AC P15205; Q62958; Q9ER21; Q9QW92;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
DE light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=96257242; PubMed=8666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
RT associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).

```

[2]  
 RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Glial tumor;  
 RX MEDLINE=92347374; PubMed=1639092;  
 RA Zauener W., Kratz J., Staunton J., Feick P., Wiche G.;  
 RT "Identification of two distinct microtubule binding domains on  
 recombinate rat MAP 1B.";  
 RL Eur. J. Cell Biol. 57:66-74 (1992).  
 [3]  
 RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.  
 RC TISSUE=Spinal cord;  
 RX MEDLINE=90059871; PubMed=2555150;  
 RA Rientz A., Grennington G., Hermans-Borgmeyer I., Kirsch J.,  
 Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;  
 RT "Neuraxin, a novel putative structural protein of the rat central  
 nervous system that is immunologically related to microtubule-  
 associated protein 5.";  
 RL EMBO J. 8:2879-2888 (1989).  
 [4]  
 RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.  
 RX MEDLINE=97405699; PubMed=9260743;  
 RA Ma D., Nothias F., Boyne L.J., Fischer I.;  
 RT "Differential regulation of microtubule-associated protein 1B (MAP1B)  
 in rat CNS and PNS during development.";  
 RL J. Neurosci. Res. 49:319-332 (1997).  
 CC -!- FUNCTION: The function of brain MAPs is essentially unknown.  
 CC Phosphorylated MAP1B may play a role in the cytoskeletal changes  
 CC that accompany neurite extension. Possibly MAP1B binds to at least  
 CC two tubulin subunits in the polymer, and this bridging of subunits  
 CC might be involved in nucleating microtubule polymerization and in  
 CC stabilizing microtubules.  
 CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate  
 CC with MAP1A and MAP1B proteins.  
 CC -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,  
 CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,  
 CC heart or muscle.  
 CC -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic  
 CC nerve levels are high early in development but decrease during  
 CC postnatal development and are low in adults. In dorsal root  
 CC ganglia levels remain high throughout development.  
 CC -!- INDUCTION: By nerve growth factor.  
 CC -!- DOMAIN: Has a highly basic region with many copies of the sequence  
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is  
 CC responsible for the binding of MAP1B to microtubules.  
 CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated  
 CC from MAP1B by proteolytic processing. It is free to associate with  
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region  
 CC of MAP1B (by similarity).  
 CC -!- PTM: Phosphorylated.  
 CC -!- SIMILARITY: TO MAP1A.  
 CC -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to  
 CC 2459) was originally described as neuraxin in Ref.3.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; U52950; AAB17068.1; -;  
 CC EMBL; X60370; CAC16162.1; -;  
 CC EMBL; X16623; CAC34620.1; ALT\_SEQ.  
 CC PIR; A56577; A56577.  
 CC InterPro; IPR000102; MAP1B neuraxin.  
 CC Fram; PF00414; MAP1B neuraxin; 10.  
 CC PROSITE; PS00230; MAP1B NEURAXIN; 8.  
 CC Microtubules; Repeat; Phosphorylation.  
 KW CHAIN ? 2459 MAP1 LIGHT CHAIN LC1.  
 FT REPEAT 1869 1885 MAP1B 1.  
 FT REPEAT 1886 1902 MAP1B 2.  
 FT REPEAT 1903 1919 MAP1B 3.

FT REPEAT 1920 1936 MAP1B 4.  
 FT REPEAT 1937 1953 MAP1B 5.  
 FT REPEAT 1954 1970 MAP1B 6.  
 FT REPEAT 1988 2004 MAP1B 7.  
 FT REPEAT 2005 2021 MAP1B 8.  
 FT REPEAT 2022 2038 MAP1B 9.  
 FT REPEAT 2039 2055 MAP1B 10.  
 FT DOMAIN 559 1035 GLU-RICH.  
 FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
 KKEE AND KKEI/V REPEATS).  
 FT DOMAIN 2224 2312 LYS-RICH.  
 FT CONFLICT 127 127 M -> V (IN REF. 1).  
 FT CONFLICT 140 140 T -> S (IN REF. 1).  
 FT CONFLICT 2112 2112 R -> K (IN REF. 3).  
 FT CONFLICT 2169 2169 L -> I (IN REF. 3).  
 SQ SEQUENCE 2459 AA; 269497 MW; 2E3F6872DEDB8A2 CRC64;  
 Query Match 11.0%; Score 82.5; DB 1; Length 2459;  
 Best Local Similarity 25.4%; Pred. No. 1.5e+02;  
 Matches 35; Conservative 25; Mismatches 49; Indels 29; Gaps 6;  
 QY 10 GKEMSTTVSBEEDFILPVYKGELEKGYQFDGWEISGFEGKDKDAGYVNLNLSKDTFTKVPVK 69  
 DB 593 GKVESKPSVTEK-----VPSKEEQSPVKAEEVAEKAATESKP-----KVTDKVVKKEIK 642  
 QY 70 -KIEKKKEEENKPTFDVSKKQNDQVNHSQLNESHKEDLQREHSQSDSTKDTVTATVL 128  
 DB 643 TKPEEKKEE--KPKKEVAKED-----KTPKKDEKPKKEEAKKEIK 686  
 QY 129 --DKNNISSKSTNNPNK 144  
 DB 687 KEEKKELKKEVKETPLK 704  
 RESULT 27  
 YMB4 YEAST  
 ID YMB4 YEAST STANDARD; PRT; 279 AA.  
 AC P49957;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Hypothetical 32.4 kDa protein in TAF40-ERV25 intergenic region.  
 GN YML014W OR YN571.04.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX PubMed=9169872;  
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,  
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RT XIII.";  
 RL Nature 387:90-93 (1997).  
 CC -!- SIMILARITY: TO S.POMBE SPAC13D6.03C AND SOME, TO C.ELEGANS  
 CC C14B1.5.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; Z49810; CAA89938.1; -;  
 CC PIR; S55105; S55105.  
 CC SGD; S0004476; YML014W.  
 CC InterPro; IPR001601; Methyltransf.



```
DR InterPro; IPR000051; SAM_bind.
KW Hypothetical protein.
SQ SEQUENCE 279 AA; 32438 MW; 2F8795BB3C640D3B CRC64;

Query Match
Best Local Similarity 11.0%; Score 82; DB 1; Length 279;
Matches 24; Conservative 14; Mismatches 39; Indels 6; Gaps 3;

QY 41 WEISFEGKQAGYVILNSKDTFFIKPVFKKIEEKEENKPTFDVSKKDNQVNHSLN 100
DB 145 WALE--QSSRRGYHEGMEQDFVFWLPSKSKPKTKSTPP--AKVTRPKPLMLNIP 199

QY 101 ESHRKEDLQ-EEHQSQSDSTKD 122
DB 200 PKRSEYLQRWKEEQQRSKSLDD 222

RESULT 28
YKTS_YEAST
ID YKTS_YEAST STANDARD; PRT; 427 AA.
AC P36046;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Hypothetical 47.4 kDa protein in PAS1-MST1 intergenic region.
GN YKL195W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RA Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L12410; AAA71968.1; -.
DR EMBL; AE010168; AAL80574.1; -.
DR HSSP; P06201; ILGR.
DR InterPro; IPR001691; GLN synth.
DR InterPro; IPR004809; GlnA.
DR InterPro; IPR001637; GlnA_adenylcn.
DR Pfam; PF00120; gln-synt; 1.
DR Pfam; PF03951; gln-synt N; 1.
DR ProDom; PD001057; Gln synt C; 1.
DR TIGRFAMs; TIGR00653; GlnA_1.
DR PROSITE; PS00180; GlnA_1; 1.
DR PROSITE; PS00181; GlnA_ATP; 1.
KW Ligase; Complete proteome.
FT BINDING 358
FT CONFLICT 203 203
FT CONFLICT 232 232
FT CONFLICT 351 351
FT CONFLICT 370 370
FT CONFLICT 376 376
FT CONFLICT 382 382
FT CONFLICT 397 397
FT CONFLICT 421 426
SQ SEQUENCE 439 AA; 50183 MW; 96563A8E9A0E0892 CRC64;

AMP (UNDER CONDITIONS OF ABUNDANT
GLUTAMINE) (BY SIMILARITY).
K -> G (IN REF. 1).
M -> L (IN REF. 1).
A -> S (IN REF. 1).
I -> L (IN REF. 1).
S -> N (IN REF. 1).
E -> G (IN REF. 1).
A -> S (IN REF. 1).
MPKOTK -> IPPDTE (IN REF. 1).

Query Match
Best Local Similarity 11.0%; Score 82; DB 1; Length 439;
Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

QY 35 GYQFDGWEISFEGKQAGYVILNSKDTFFI-----KP-----VFK 69
DB 45 GISFDGSSVPGFGIETSDLVFKADPDYVEVPDVARVYGFYVKNPKYCADPRGILK 104
QY 70 KIEEKKEEN-----KPTFDVSKKDN----- 91
DB 105 RALELEKEGKYKAYIGPEPEFYLPKNGTWELEIPDVGGYFDILTLDKARIRRIAEYM 164
QY 92 -----POVNHSQLNESHKEDLQREHSQKSD---STKDVTTATVLDKNNI 133
DB 165 PSFGLIPEVLHVEGVKAQHEIDFRYDEALKADNIVSVFKYITKVAEMHGL 215

RESULT 30
```

```
DR InterPro; IPR000051; SAM_bind.
KW Hypothetical protein.
SQ SEQUENCE 279 AA; 32438 MW; 2F8795BB3C640D3B CRC64;

Query Match
Best Local Similarity 11.0%; Score 82; DB 1; Length 279;
Matches 24; Conservative 14; Mismatches 39; Indels 6; Gaps 3;

QY 41 WEISFEGKQAGYVILNSKDTFFIKPVFKKIEEKEENKPTFDVSKKDNQVNHSLN 100
DB 145 WALE--QSSRRGYHEGMEQDFVFWLPSKSKPKTKSTPP--AKVTRPKPLMLNIP 199

QY 101 ESHRKEDLQ-EEHQSQSDSTKD 122
DB 200 PKRSEYLQRWKEEQQRSKSLDD 222

RESULT 28
YKTS_YEAST
ID YKTS_YEAST STANDARD; PRT; 427 AA.
AC P36046;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Hypothetical 47.4 kDa protein in PAS1-MST1 intergenic region.
GN YKL195W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RA Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z28195; CAA82039.1; -.
DR PIR; S38032; S38032
DR SGD; S0001678; YKL195W.
KW Hypothetical protein.
SQ SEQUENCE 427 AA; 47416 MW; 44386D250DE5DED4 CRC64;

Query Match
Best Local Similarity 11.0%; Score 82; DB 1; Length 427;
Matches 29; Conservative 18; Mismatches 43; Indels 16; Gaps 3;

QY 31 ELEKGQFDGWEISFEGKQAGYVILNSKDTFFIKPVFKKIEEKEENKPTFDVSKKD 90
DB 142 ETEAGPOLGGDKIGASKVAEDGELVLAEDN-----KSSDKDTSK---VSTKDD 191
QY 91 NPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISK 136
DB 192 -----EQSNEDNATANNQKDNISENSESTSDKTLDNAGSSE 231

RESULT 29
GLNA_PYRUFU
ID GLNA_PYRUFU STANDARD; PRT; 439 AA.
AC Q05907;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
OS Pyrococcus furiosus.
```

01-FEB-1995 (Rel. 31, Last sequence update)  
28-FEB-2003 (Rel. 41, Last annotation update)  
Penicillin-binding protein 1A/1B (PBPI) [includes: Penicillin-  
insensitive transglycosylase (EC 2.4.2.-) (Peptidoglycan TGase);  
Penicillin-sensitive transpeptidase (EC 3.4.-.-) (DD-transpeptidase)]  
PONA.  
Bacillus subtilis.  
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
NCBI\_TaxID=1423;  
[1]  
SEQUENCE FROM N.A., AND SEQUENCE OF 499-515.  
STRAIN=168;  
MEDLINE=95i113769; PubMed=7814321;  
Popham D.L., Setlow P.;  
"Cloning, nucleotide sequence, and mutagenesis of the Bacillus  
subtilis ponA operon, which codes for penicillin-binding protein  
(PBPI) 1 and a pap-related factor.";  
J. Bacteriol. 177:326-335(1995).  
[2]  
SEQUENCE FROM N.A.  
STRAIN=168 / Marburg;  
MEDLINE=963349105; PubMed=8760912;  
Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,  
Serror P.;  
"Sequence analysis of the Bacillus subtilis chromosome region between  
the serA and kds loci cloned in a yeast artificial chromosome.";  
Microbiology 143:2005-2016(1996).  
[3]  
SEQUENCE FROM N.A.  
STRAIN=168;  
MEDLINE=980444033; PubMed=9384377;  
Kunat F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,  
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
Choi S.-K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.I.,  
Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,  
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
Guiseppi G., Guy B.J., Haga K., Hatech J., Harwood C.R., Henaut A.,  
Hilbert H., Holgappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
Joris B., Karamata D., Kasahara Y., Klaer-Blandhard M., Klein C.,  
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
Parro V., Pohl T.M., Portetelle D., Porwoll S., Prescott A.M.,  
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
Rieger M., Rivalto C., Roche E., Roche B., Rose M., Sadate Y.,  
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,  
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
Viari A., Wambut R., Wedler H., Wedler H., Weitzenecker T.,  
Winters P., Wipst A., Yamamoto H., Yanane K., Yasumoto K., Yata K.,  
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
"The complete genome sequence of the gram-positive bacterium Bacillus  
subtilis";  
Nature 390:249-256(1997).  
[4]  
GROWTH REQUIREMENTS.  
STRAIN=168;  
MEDLINE=98389671; PubMed=9721295;  
Murray T., Popham D.L., Setlow P.;  
"Bacillus subtilis cells lacking penicillin-binding protein 1 require  
increased levels of divalent cations for growth.";  
J. Bacteriol. 180:4555-4563(1998).  
[5]  
SUBCELLULAR LOCATION.  
STRAIN=168;



```
SQ SEQUENCE 949 AA; 105961 MW; FA8969B0C64B3278 CRC64;
Query Match 11.0%; Score 82; DB 1; Length 949;
Best Local Similarity 28.7%; Pred. No. 58;
Matches 29; Conservative 20; Mismatches 20; Indels 32; Gaps 5;
QY 60 KDTFTKPVKKTEEEKKEENKTFDVSKKD-----NPK-VNHSQLNES 102
DB 98 EETKQPKPKTKKKEKAPAP---IHKKEIEIVNTFENPTPLVNTPKAVSHSQIEKA 154
QY 103 HKKEDLOREHSKSDSTKDVATVLDKNNISKSTNNPN 143
DB 155 --KQKLEIQKREA-----LNKLTQNTNTNNAN 183
RESULT 33
LMD1 HUMAN STANDARD; PRT; 572 AA.
AC P29536;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Leiomodlin 1 (Leiomodlin, muscle form) (64 kDa autoantigen D1) (64 kDa
DE autoantigen ID) (64 kDa autoantigen ID3) (Thyroid-associated
DE ophthalmopathy autoantigen) (Smooth muscle leiomodlin) (SM-Lmod).
GN LMOD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=91225220; PubMed=206759;
RA Dong Q., Ludgate M., Vaseart G.;
RT "Cloning and sequencing of a novel 64-kDa autoantigen recognized by
RT patients with autoimmune thyroid disease.";
RL J. Clin. Endocrinol. Metab. 72:1375-1381(1991).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=99451105; PubMed=10520227;
RA Conley C.A., Fowler V.M.;
RT "Localization of the human 64kD autoantigen D1 to myofibrils in a
RT subset of extraocular muscle fibers.";
RL Curr. Eye Res. 19:313-322(1999).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=21248187; PubMed=11350761;
RA Conley C.A.;
RT "Leiomodlin and tropomodulin in smooth muscle.";
RL Am. J. Physiol. 280:C1645-C1656(2001).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=21218919; PubMed=11318603;
RA Conley C.A., Fritz-Six K.L., Almenar-Queralt A., Fowler V.M.;
RT "Leiomodlins: larger members of the tropomodulin (Tmod) gene family.";
RL Genomics 73:127-139(2001).
CC -1- SUBCELLULAR LOCATION: CYTOSKELETON.
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE (HEART, SKELETAL MUSCLE, COLON
CC AND SMALL INTESTINE), A SUBSET OF STRIATED MUSCLE FIBERS, AND AT
CC LOW LEVEL IN THYROID.
CC -1- DISEASE: RECOGNIZED BY PATIENTS WITH AUTOIMMUNE THYROID DISEASE.
CC -1- SIMILARITY: BELONGS TO THE TROPOMODULIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X54162; CAA38101.1; -.
```

```
DR PIR; S18732; S18732.
DR Genew; HGNC:6647; LMOD1.
DR MIM; 602715; -.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0009405; P:pathogenesis; TAS.
DR InterPro; IPR004934; Tropomodulin.
DR InterPro; IPR003124; WH2.
DR Pfam; PF03250; Tropomodulin; 1.
DR Pfam; PF02205; WH2; 1.
DR SMART; SM00246; WH2; 1.
DR Antigen; Repeat; Cytoskeleton.
KW DOMAIN 137 265 8 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 137 152 1.
FT REPEAT 153 168 2.
FT REPEAT 169 184 3.
FT REPEAT 185 200 4.
FT REPEAT 201 216 5.
FT REPEAT 217 232 6.
FT REPEAT 233 248 7.
FT REPEAT 249 265 8.
FT DOMAIN 480 499 5 X 4 AA APPROXIMATE TANDEM REPEATS.
SQ SEQUENCE 572 AA; 63737 MW; D4B42F8E0523DE94 CRC64;
Query Match 10.9%; Score 81.5; DB 1; Length 572;
Best Local Similarity 22.8%; Pred. No. 37;
Matches 38; Conservative 21; Mismatches 57; Indels 51; Gaps 6;
QY 29 KGELEKGFQFDGWEISGFEGK--KDAGYVNLKSDTFIKVPFK-----IEEK 74
DB 99 RGLKKSFSRDRDEAGGKSGEKPEKIIRGIDKGRVRAAVDKKAGDKGRGEERAVATK 158
QY 75 KEENK-----PTFDVSKKNDPQVNHSQLNESHKE----- 106
DB 159 KEEKKGGDRNTGLSRDKDKREEMKEVAKDEDEKVGGERNTDTRKEGEMKRAAGNT 218
QY 107 DLQREHSQK-----SDSTKDVAT-----VLDKNNISKSTNNPNK 144
DB 219 DMKDEKVKRGNTGNTDKDDEKVKKNEPLHEKEAKDDSKTKTPEK 265
RESULT 34
DRSL YEAST STANDARD; PRT; 752 AA.
ID DRSL YEAST
AC P32892;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable ATP-dependent RNA helicase DRS1.
GN DRS1 OR YLL008W OR L1345.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93087480; PubMed=1454790;
RA Ripmaster T.L., Vaughn G.P., Woolford J.L. Jr.;
RT "A putative ATP-dependent RNA helicase involved in Saccharomyces
RT cerevisiae ribosome assembly.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:11131-11135(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / FY23;
RX MEDLINE=96405918; PubMed=8810043;
RA Miosga T., Zimmermann F.K.;
RT "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
RT a 43.7 kb fragment of chromosome XII including an open reading frame
RT homologous to the human cystic fibrosis transmembrane conductance
RT regulator protein CFTR.";
RL Yeast 12:693-708(1996).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
```

```

RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Ansoerge W.,
RA Benes V., Brueckner M., Dubois H., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koettler P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Mioega T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestarazu L.A., Vandenbol M., Verhasselt P.,
RA Vierdeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Honeisel J.D.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:87-90(1997).
CC -!- FUNCTION: PROBABLE HELICASE INVOLVED IN RIBOSOME ASSEMBLY.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; L00683; AAA34666.1; -
CC EMBL; X91488; CA62783.1; -
CC EMBL; Z73113; CA97452.1; -
CC EMBL; PIR; S64750; S64750.
CC HSSP; Q58083; 1HV8.
CC SGD; S0003931; DRS1.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR000629; DEAD_box.
CC InterPro; IPR001650; Helicase_C.
CC Pfam; PF00270; DEAD; 1.
CC Pfam; PF00271; Helicase_C; 1.
CC SMART; SM00487; HEDC; 1.
CC SMART; SM00490; HELIC; 1.
CC PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
KW ATP-binding; RNA-binding; Helicase; Nuclear protein.
FT DOMAIN 170 190 POLY-GLU.
FT NP BIND 275 282 ATP (POTENTIAL).
FT SITE 385 388 DEAD BOX.
FT CONFLICT 1 68 MVCTGKYSLNLDVFPTLSDSDVPLDSSDDEKVEAKKT
FT KRRGKGNKKYKVSQNLNDEVDHEDLD -> MTKSLRLRL
FT RSRGRVRLTRKRLVKGITSMFPMKWT (IN REF. 1).
SQ SEQUENCE 752 AA; 84843 MW; 60747607A6E5E4A8 CRC64;
Query Match 10.9%; Score 81.5; DB 1; Length 752;
Best Local Similarity 21.0%; Pred. No. 50;
Matches 30; Conservative 26; Mismatches 48; Indels 39; Gaps 6;
Qy 18 VSEEDFLPVYVYKLEKGYQFD-----GWEI--SGFEKGKDGAGYVNLKSDTFI 64
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 VSEGNLDEVDHEDLDAGFKFDLADDTTNSFQGNFLAEGSNKDDAEAFV--KKDVL 110
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 65 KPVFKK-----IEKKEENKPTFDVSKKDN-----PQVHNSQLNE 101
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 DKIRRRKGLVMAHIDSKQEEETEKE-KVEKENDSDDELDAMDFGGMNNGENQS 169
102 SHRKEDLQREHSQKSDSTKDV 124
Db 170 EEEEEEKKEEEEEEEQEEEMT 192
RESULT 35
DACA BACSU
ID DACA BACSU STANDARD; PRT; 443 AA.
AC P08750;
DC 01-AUG-1988 (Rel. 08, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

```

DE D-alanyl-D-alanine carboxypeptidase precursor (EC 3.4.16.4) (DP-
DE peptidase) (DD-carboxypeptidase) (CPase) (PBPs).
GN DACA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OK NCBI_TaxID=1423;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RL subtilis chromosome containing the replication origin.";
RN DNA Res. 1:1-14(1994).
RN [2]
RN SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaier-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [3]
RN SEQUENCE OF 32-102.
RX MEDLINE=80182289; PubMed=6768745;
RA Waxman D.J., Strominger J.L.;
RT "Sequence of active site peptides from the penicillin-sensitive D-
RT alanine carboxypeptidase of Bacillus subtilis. Mechanism of
RT penicillin action and sequence homology to beta-lactamases.";
RL J. Biol. Chem. 255:3964-3976(1980).
RN [4]
RN SEQUENCE OF 103-443 FROM N.A.
RX MEDLINE=86250602; PubMed=3087956;
RA Todd J.A., Roberts A.N., Johnstone K., Piggot P.J., Winter G.,
RA Ellar D.J.;
RT "Reduced heat resistance of mutant spores after cloning and
RT mutagenesis of the Bacillus subtilis gene encoding penicillin-binding
RT protein S.";
RL J. Bacteriol. 167:257-264(1986).
RN [5]
RN SEQUENCE OF 414-443.
RX MEDLINE=81117303; PubMed=6780559;
RA Waxman D.J., Strominger J.L.;
RT "Primary structure of the COOH-terminal membranous segment of a
RT penicillin-sensitive enzyme purified from two Bacilli.";
RL J. Biol. Chem. 256:2067-2077(1981).

```





```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59720; CAA42405.1; -
DR EMBL; X53998; -; NOT_ANNOTATED_CDS.
DR PIR; S74279; S74279.
DR SGD; S0000566; YCL061C.
DR GO; GO:0006347; P:chromatin silencing at HML and HMR (sensu S. . .); IGI.
DR GO; GO:0006348; P:chromatin silencing at telomere; IGI.
DR GO; GO:0000076; P:DNA replication checkpoint; IGI.
KW Hypothetical protein.
FT CONFLICT 505 505 L -> V (IN REF. 3).
FT CONFLICT 567 567 MISSING (IN REF. 3).
SQ SEQUENCE 853 AA; 97946 MW; 16B09FCC0BF248D1 CRC64;
-----
Query Match 10.8%; Score 81; DB 1; Length 853;
Best Local Similarity 21.7%; Pred. No. 62;
Matches 34; Conservative 21; Mismatches 68; Indels 34; Gaps 5;

QY 11 KEMSTIVSEEDFILPVY-----KGELEKGYQFDGWEISGFEG-----KKD 51
DB 293 KEKREKLEENDFQNAHDSGSDSGSFGALSGNEIADYESSGSENDNRSEDSKED 342

QY 52 AGYVNLKDTPIKPVFKIEKKEENKPTPDVSKKXNDPQVNHSHQSHKEDLORE 111
DB 343 DEILKQKKSHVHHIINESDTEVEAKP-----KEKADESLPKRIALNIGHYDNI--- 395

QY 112 EHSQKSDTKDVTATVLDKNNI-----SSKSTNNPNK 144
DB 396 ----GEDTDKFOETVNLDTQNIIEVMAERTNIENEVK 428

RESULT 38
YAE6 SCHPO
ID YAE6 SCHPO STANDARD; PRT; 1325 AA.
AC Q09847;
DT 01-FEB-1996 (Rel. 33.; Created)
DT 01-FEB-1996 (Rel. 33.; Last sequence update)
DT 28-FEB-2003 (Rel. 41.; Last annotation update)
DE Hypothetical protein C23D3.06c in chromosome I.
GN SPAC23D3.06C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Brooks J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Gough J., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

```

```

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.K., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z64354; CAA91241.1; -
DR PIR; S62497; S62497.
DR GenedB SPombe; SPAC23D3.06C; -.
DR Pfam; PF03093; Nucleoporin_FG; 10.
KW Hypothetical protein.
SQ SEQUENCE 1325 AA; 145776 MW; EEF9D52FB7F0E68D CRC64;
-----
Query Match 10.8%; Score 81; DB 1; Length 1325;
Best Local Similarity 23.3%; Pred. No. 99;
Matches 35; Conservative 19; Mismatches 40; Indels 56; Gaps 6;

QY 41 WEISGFEGKKDAGYVNLKSD-----TFIKPVFKIEKKEEN---KP-TFDV 85
DB 685 WGASTFQSKPQPSFSGFLTLDDKSNTPGKNSIFGKTAETQVEQKPPNNVLTKEPFPAP 744

QY 86 SKK-----KDNQVNHSHQSHKEDLOREHSOKSDSKDVTVA 125
DB 745 SDKSMFAANIPISAGEGLDQKTSKALPSTGTTKUSEND-----NEKABESNEIKGFNT 797

QY 126 TVLDKNNISKS-----TTNN 141
DB 798 TIARQNDKSKSEKSGKASVANNMALKNSTNN 827

RESULT 39
UN89 CAEBL
ID UN89 CAEBL STANDARD; PRT; 6632 AA.
AC Q01761; Q17362;
DT 15-SEP-2003 (Rel. 42.; Created)
DT 15-SEP-2003 (Rel. 42.; Last sequence update)
DT 15-SEP-2003 (Rel. 42.; Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT assembly, encodes a giant modular protein composed of Ig and signal
RT transduction domains."
RL J. Cell Biol. 132:835-848(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Du Z., Le T.T., Wilson R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

```







**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 25.4955 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_630\_773

Perfect score: 748

Sequence: 1 HRVTVTIQKMSSTIVSE.....ATVLKKNISSKSTNNPNK 144

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*\*

- 1: sp\_archaea:\*\*
- 2: sp\_bacteria:\*\*
- 3: sp\_fungi:\*\*
- 4: sp\_human:\*\*
- 5: sp\_invertebrate:\*\*
- 6: sp\_mammal:\*\*
- 7: sp\_mhc:\*\*
- 8: sp\_organelle:\*\*
- 9: sp\_phase:\*\*
- 10: sp\_plant:\*\*
- 11: sp\_rodent:\*\*
- 12: sp\_virus:\*\*
- 13: sp\_vertebrate:\*\*
- 14: sp\_unclassified:\*\*
- 15: sp\_rvirus:\*\*
- 16: sp\_bacteriap:\*\*
- 17: sp\_archaeap:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	748	100.0	2119	2	Q9AHT5
2	748	100.0	2140	16	Q97RY6
3	745	99.6	2144	16	Q8DQP7
4	744	99.5	2144	2	Q9S4M8
5	112	15.0	361	5	Q95PI5
6	110	14.7	346	5	Q9U0G0
7	110	14.7	379	5	Q25705
8	110	14.7	775	16	Q8CPK8
9	109	14.6	3008	5	Q8I436
10	108.5	14.5	600	5	Q77355
11	108.5	14.5	1038	13	Q90784
12	108	14.4	354	5	Q25995
13	108	14.4	354	5	Q8IJ55
14	107	14.3	829	5	Q8IF53
15	106.5	14.2	470	10	Q9FJK9
16	104.5	14.0	379	5	Q9U6C4

17	104.5	14.0	380	5	Q26019
18	104	13.9	951	5	Q96229
19	103.5	13.8	325	5	Q44016
20	103.5	13.8	379	5	Q25706
21	102.5	13.7	2081	10	Q9LH98
22	102.5	13.6	382	5	Q9V7U0
23	101.5	13.6	556	5	Q9V7I9
24	101.5	13.6	785	5	Q9GQ82
25	101	13.5	329	5	Q9NFV9
26	100	13.4	312	16	Q9PPL5
27	100	13.4	1130	5	Q8IJ24
28	99	13.2	211	5	Q9I488
29	99	13.2	1859	5	Q8IC27
30	99	13.2	2563	5	Q8I3A0
31	98	13.1	157	5	Q9VQV0
32	97.5	13.0	449	5	Q8IHW3
33	97.5	13.0	556	5	Q9S593
34	97	13.0	558	5	Q8I5T1
35	97	13.0	988	10	Q9STN4
36	96	12.8	238	5	Q8I226
37	96	12.8	531	5	Q8I426
38	96	12.8	1795	5	Q8IE35
39	95.5	12.8	375	4	Q14712
40	95.5	12.8	635	16	Q8ESX9
41	95.5	12.8	641	10	Q93ZQ2
42	95.5	12.8	644	10	Q9M1D2
43	95	12.7	535	5	Q17595
44	95	12.7	622	16	Q980A1
45	95	12.7	3026	5	Q8ILS9

#### ALIGNMENTS

#### RESULT 1

Q9AHT5 PRELIMINARY; PRT; 2119 AA.  
AC Q9AHT5;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Serine protease (Fragment).  
GN PRTA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_taxid=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
RA Gayle A., Brewah Y.A., Walsh W., Warren P., Lathigra R., Hanson M.,  
RA Langermann S., Johnson S., Koenig S.;  
RT "Use of a Whole Genome Approach To Identify Vaccine Molecules  
Affording Protection against Streptococcus pneumoniae Infection.";  
RL Infect. Immun. 69:1593-1598(2001).  
CC -|- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
AN AMIDE BOND (BY SIMILARITY).  
DR EMBL; AF291699; AAK19159.1; -;  
DR HSSP; P00782; 2S8T.  
DR MEROPS; S08.064; -;  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00745; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRfams; TIGR01167; LPXTG\_anchor; 1.

```
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Cell wall; Peptidoglycan-anchor; Protease.
FT NON_TER 1
SQ SEQUENCE 2119 AA; 238226 MW; 517F9B7F6B960A6A CRC64;

Query Match      100.0%; Score 748; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 9.4e-49;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVNLK 60
Db 1942 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVNLK 2001

QY 61 DTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHSHKSDST 120
Db 2002 DTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHSHKSDST 2061

QY 121 KDVTATVLDKNNISSKSTNNPNK 144
Db 2062 KDVTATVLDKNNISSKSTNNPNK 2085

RESULT 2
Q97RY6 PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Serine protease, subtilase family.
GN SP0641.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FIG4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umavam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapfel E., Khouri H., Wolf A.M., Utterback T.R., Hansen, C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae";
RL Science 293:498-506 (2001).
DR EMBL; AE007373; AAK74791.1; -.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRAME; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Protease; Complete proteome.
```

```

SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;

Query Match      100.0%; Score 748; DB 16; Length 2140;
Best Local Similarity 100.0%; Pred. No. 9.5e-49;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVNLK 60
Db 1963 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVNLK 2022

QY 61 DTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHSHKSDST 120
Db 2023 DTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHSHKSDST 2082

QY 121 KDVTATVLDKNNISSKSTNNPNK 144
Db 2083 KDVTATVLDKNNISSKSTNNPNK 2106

RESULT 3
Q8DQP7 PRELIMINARY; PRT; 2144 AA.
AC Q8DQP7;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN PrtA OR SP0561.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., BURGESS S.,
RA DeHoff R.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore B., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., Mcleaster K., Mundy C.W., Niclas T.I.,
RA Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rocky P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717 (2001).
DR EMBL; AE008434; AAK9365.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match      99.6%; Score 745; DB 16; Length 2144;
Best Local Similarity 99.3%; Pred. No. 1.6e-48;
Matches 143; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVNLK 60
Db 1967 HRTVTIIONGKMSSTIVSEEDFILPVYKGELEKGYQFDGWISGFEKGDAGYVNLK 2026

QY 61 DTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHSHKSDST 120
Db 2027 DTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKREDLQREHSHKSDST 2086

QY 121 KDVTATVLDKNNISSKSTNNPNK 144
Db 2087 KDVTATVLDKNNISSKSTNNPNK 2110

RESULT 4
Q9S4M8 PRELIMINARY; PRT; 2144 AA.
AC Q9S4M8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
```

01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
Cell wall-associated serine proteinase PrtA precursor.  
PRTA.  
Streptococcus pneumoniae.  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
NCBI\_TaxID=1313;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=3 B;  
Bethe G., ten Thoren E., Bongarts R.J.M., Heinz H.-P., Zysk G.;  
"Cloning and sequencing of a novel surface protease of Streptococcus  
pneumoniae";  
RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RL SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
CC -!-  
DR EMBL; AF127143; AAD48399.1; -.  
DR HSP; P00782; 2SBI.  
DR MEROPS; S08.064; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF02225; PA; 1.  
DR PRINTS; PF00882; Peptidase\_S8; 2.  
DR PRINTS; PF00723; SUBTILISIN.  
DR TIGRfams; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS; 1.  
KW Cell wall; Peptidoglycan-anchor; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 2144 CELL WALL-ASSOCIATED SERINE PROTEINASE  
FT PRTA.  
SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;  
  
Query Match 99.5%; Score 744; DB 2; Length 2144;  
Best Local Similarity 98.6%; Pred. No. 1.9e-48;  
Matches 142; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 HRVTVIQNGEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAVVNLK 60  
Db 1967 HRVTVIQNGEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAVVNLK 2026  
  
Qy 61 DTFIKPVFKIEEKEENKPTFDVSKKKNPQVNHSQLNESHKREDLQREHSQKSDST 120  
Db 2027 DTFIKPVFKIEEKEENKPTFDVSKKKNPQVNHSQLNESHKREDLQREHSQKSDST 2086  
  
Qy 121 KDVTATVLDKNNISSKSTTNNPK 144  
Db 2087 KDVTATVLDKNNISSKSTTNNPK 2110  
  
RESULT 5  
Q95PI5 ID Q95PI5 PRELIMINARY; PRT; 361 AA.  
AC Q95PI5  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Merozoite surface protein 3 (fragment).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVO;  
RA Hiseida H., Saul A., Long C.A., Miller L.H., Stowers A.W.;  
RT "Merozoite Surface Protein 3 and Protection Against Malaria in Aotus

Monkeys";  
RT Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY044180; AAK94780.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 361 361  
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;  
  
Query Match 15.0%; Score 112; DB 5; Length 361;  
Best Local Similarity 23.9%; Pred. No. 0.57;  
Matches 37; Conservative 28; Mismatches 48; Indels 42; Gaps 6;  
  
Qy 28 YKGELEKGYQ-----FD---GWEISGF--EGKKDAG-----YVI 56  
Db 165 YAGKVKDYERAKNAVQKANQAVLKAKEASSYDILGWFEFGGVPHEKKEENMLSHLYVS 224  
Qy 57 NLSKDTPIKPVFKIEEKEENKPTFDVSKKKNPQVNHSQLNESHKREDLQREHSQK 116  
Db 225 SKKENISKENDVDLDE-KEEEAETEERLEEKNEETEISEDEDEEEEEEEEEEE 283  
  
Qy 117 SDSTKQVTVLDKN-----NISKSTNN 141  
Db 284 NDKKKEQESQNNENNNDQKDMQAQLISKNNNN 318  
  
RESULT 6  
Q9U0G0 ID Q9U0G0 PRELIMINARY; PRT; 346 AA.  
AC Q9U0G0  
DT 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Merozoite surface protein 3 (fragment).  
OS MSP3.  
GN Plasmodium reichenowi.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5854;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20416497; PubMed=10960178;  
RA Okenu D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in  
Plasmodium reichenowi and Plasmodium falciparum";  
RL Mol. Biochem. Parasitol. 109:185-188 (2000).  
DR EMBL; AJ252286; CAB65754.1; -.  
FT NON\_TER 1 1  
FT NON\_TER 346 346  
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAPA010 CRC64;  
  
Query Match 14.7%; Score 110; DB 5; Length 346;  
Best Local Similarity 26.2%; Pred. No. 0.78;  
Matches 34; Conservative 26; Mismatches 54; Indels 16; Gaps 5;  
  
Qy 27 VYKGELEKGYQFD-GWEISGF--EGKKDAG-----YVINLSKDTFIKPVFKIEEKEE 78  
Db 184 VLKAKEASSYNYILGWFEFGGVPHEKKEENMLSHLYSSKDKKENISKENDVDLDE-KEE 242  
Qy 79 NKPTFDVSKKKNPQVNHSQLNESHKREDLQREHSQKSDSTKQVTV-----LDKN 131  
Db 243 AETGGELEKNEETEISEINEDEEEEEEEEEEEENNKKEQAKESQNDQKEDMEAQ 302  
  
132 NISKSTNN 141  
303 NLISKNNNN 312  
  
RESULT 7  
Q25705 ID Q25705 PRELIMINARY; PRT; 379 AA.  
AC Q25705  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
DE Polymorphic antigen.

```
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RX MEDLINE=98156743; PubMed=9497029;
RA McCall D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
  Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL: U08851; AAC4781.1;
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

Query Match 14.7%; Score 110; DB 5; Length 379;
Best Local Similarity 23.4%; Pred. No. 0.86;
Matches 37; Conservative 30; Mismatches 41; Indels 50; Gaps 7;

QY 28 YGELEKGYQ-----FD---GWEISGF--EGKKDAG-----YVI 56
DB 185 YAGKEKDYERAKNAYQKQAVLKAKAEASSYDILGWFEFGGVPEHKKEENMLSHLYVS 244
QY 57 NLSKDTFIKVPFKIEEKEEENKPTFDVSKKDNQVNHSLN-----EGH 103
DB 245 SKDKENISKENDVLDE--KEEAETEELEKEEETESISEDEEEEEEEKEENE 303
QY 104 RKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 141
DB 304 KKKEQKEQSNENNQKDMEA-----QNLISKNNQNN 336

RESULT 8
Q8CPK8 PRELIMINARY; PRT; 775 AA.
AC Q8CPK8;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DE Penicillin-binding protein 1.
GN SE0856.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
  Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AE016746; AAC04453.1;
KW Complete proteome.
SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;

Query Match 14.7%; Score 110; DB 16; Length 775;
Best Local Similarity 27.0%; Pred. No. 1.8;
Matches 43; Conservative 22; Mismatches 52; Indels 42; Gaps 8;

QY 5 VTIQNGKMSSTIVSEEDFILPVK-----GELEKGYQFDGW---EISGFE-----G 48
DB 625 IITIGNGKQIKQOSVKSGTKVLPHSKVMLTDSGLTNP-DWTGKTREDVLAFEDLTIKIVS 683
QY 49 KKDAGYVIN--ISKDTFIKVPFKIEEKEEENKPTFDVSKKDNQVNHSLN-----KKDNQVNHSLN 102
DB 684 TKNGFVNTQSIKSGQIK-----NKDKIEVLSAEDTDDQKXTDESDSN 730
QY 103 HKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 141
DB 731 KSKKQKADHDHNTSSSTKN-----DKSNADSKNDSDD 763

RESULT 9
Q8I436 PRELIMINARY; PRT; 3008 AA.
ID Q8I436
```

```
AC Q8I436;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFE0325W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
  Hall N., Bowman S., Churcher C., Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
  Buckee C.O., Burrows C., Cherevach I., Chillingworth C., Corton C.,
  Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
  Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
  Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
  Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
  Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
  Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
  Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
  Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
  Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
  Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
  Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
  Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RL "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
DR EMBL: AL929351; CAD51431.1;
KW Hypothetical protein.
SQ SEQUENCE 3008 AA; 356023 MW; 60BCBBEE15C599B4 CRC64;

Query Match 14.6%; Score 109; DB 5; Length 3008;
Best Local Similarity 32.4%; Pred. No. 9.4;
Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

QY 54 YVINLSK---DTFIKVPFKIEEKEEENKPTFDVSKKDNQVNHSLN-----LNESHKKE 106
DB 2310 YDIELSKIEKFGASIGPVFTD-EENKKEENK--EVNKKENKKEENKKEENKKEENK 2366
QY 107 DLQREH---SOKSDSTKDVATVLDKNNISSK-----STNNPNK 144
DB 2367 ENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENK 2414

RESULT 10
O77355 PRELIMINARY; PRT; 600 AA.
AC O77355;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE Hypothetical 71.7 kDa protein.
GN PFC0465C, MAL3P4.20.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7.
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
  Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
  Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
  Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,
  Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
  Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
```

[illegible]



```
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014834; AAN35542.1; -.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 14.4%; Score 108; DB 5; Length 354;
Best Local Similarity 23.2%; Pred. No. 1.1; Indels 44; Gaps 6;
Matches 36; Conservative 29; Mismatches 46;

QY 7 IQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLS 59
DB 181 VLKAKEASS-----YDYL-----GWEFGGVPEHKKEENMLSHLVSSKD 221

QY 60 KDTFIKPVFKTIEEKGE-----ENKPTFDVSKKDNQVNHSQLNSHRKE 106
DB 222 KENISKENDVDLDEKEEAETEEBLEKNEETSEISEDEEESEEEKEEENDKKK 281

QY 107 DLQREHSQKSSTKDVTVATVLDKNNISKSTNN 141
DB 282 EQEKEQSNENNDQKDMEA-----QNLISKNNNN 311

RESULT 14
Q815F3 PRELIMINARY; PRT; 829 AA.
ID Q815F3;
AC Q815F3;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFL1275C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=2225705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Sun B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014848; AAN36341.1; -.
SQ SEQUENCE 829 AA; 98816 MW; EF2675E301B2CE93 CRC64;

Query Match 14.3%; Score 107; DB 5; Length 829;
Best Local Similarity 24.8%; Pred. No. 3.4;
Matches 36; Conservative 30; Mismatches 43; Indels 36; Gaps 6;

QY 16 TVTSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGKAGVIVLSKDTFIKPVFKTIEKK 75
DB 519 TLMTERDIVLDFHPYIMKKYHLN-----KKETLFFNLSN-----FRIEIKNK 561

QY 76 EENKPT-----FDVSKKDN--PQVNHSQLNESH--RKEDLQREHSQKSDS 119
DB 562 RUKKKGTHNNKNDAAEYMLKTKIKKKKKNFENNNTNNIKKNNKLVEH---DNS 618

QY 120 TKDVTATVLDKNNISKSTNNPNK 144
DB 619 LKQEQIINDKNVIEHTKIYDNQK 643
```

```
RESULT 15
Q9FJK9 PRELIMINARY; PRT; 470 AA.
ID Q9FJK9;
AC Q9FJK9;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Gb|AAF20218.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308(1998).
DR EMBL; AB015468; BAB10694.1; -.
SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72E35AC54 CRC64;

Query Match 14.2%; Score 106.5; DB 10; Length 470;
Best Local Similarity 20.3%; Pred. No. 2;
Matches 36; Conservative 31; Mismatches 65; Indels 45; Gaps 5;

QY 1 HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF-----GKKDA 52
DB 84 NRVTDTQVNNNGSK-----YVQLARRIYDE--EATGSAQRIDHPNQKNV 131

QY 53 GVINLSKDTFIKPVFKTIEEKKEENKPTFDVSKKKN-----PQ 93
DB 132 GITEKAPENSPIETSHVEDNKRNNQKNTAAKSENASVRSVFGADHRAEVMGKPM 191

QY 94 VNHSQLNE-----SHRKEDLQREHSQKSSTKDVTVATVLDKNNISKSTNNPNK 144
DB 192 ENRDQVRQTESAEKSHRKNVTKEKPRDQGVKKTEAKDRNKEKEEKTESINK 248

RESULT 16
Q9U6C4 PRELIMINARY; PRT; 379 AA.
ID Q9U6C4;
AC Q9U6C4;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Polymorphic antigen.
GN MSP-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCCL/HN;
RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;
RT "Sequence of Plasmodium falciparum secreted polymorphic antigen
RT gene.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF198190; AAF04099.1; -.
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1F9D5F25 CRC64;

Query Match 14.0%; Score 104.5; DB 5; Length 379;
Best Local Similarity 24.5%; Pred. No. 2.3;
Matches 38; Conservative 28; Mismatches 44; Indels 45; Gaps 7;

QY 7 IQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLS 59
DB 207 VLKAKEASS-----YDYL-----GWEFGGVPEHKKEENMLSHLVSSKD 247
```



Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;

QY 1 HRTVTIQKMSSTIVSEEDFILPVYK-GEL--EKGYQFDGWEISGFEK-----49

DB 57 HRTITSIKN--RFSVKIGDEEKLFRISKNGELIVLNELEFDNPHIK--EGKHLRKSXMF 112

QY 50 ---KDAGYV-----INLSKDTFKPV-----FKK-----70

DB 113 NHIKSGYATNNEEIEFLSCITLKEITAOIKRNSYKRNINIKLPBEEEBEEREE 172

QY 71 ---IEEKKEENKPTFDVSKKNDPQVNHSQLNSHRKE-----DLQREHSQKSDTK 121

DB 173 EEEERQEEVEKPTISEEBETPAVSEEBEEREEETPAVSEEBEEREEQEDKEK 232

QY 122 D 122

DB 233 D 233

RESULT 20

Q25706 PRELIMINARY; PRT; 379 AA.

AC Q25706;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Polymorphic antigen.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.

OX NCBI\_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CS12;

RX MEDLINE=98156743; PubMed=9497029;

RA McCall D.J., Anders R.F.;

RT "Conservation of structural motifs and antigenic diversity in the Plasmodium falciparum merozoite surface protein-3 (MSP-3).";

RL Mol. Biochem. Parasitol. 90:21-31(1997).

DR EMBL; U08852; AAC47832.1; -

SQ SEQUENCE 379 AA; 43302 MW; ABF9D54E1ED91A24 CRC64;

Query Match 13.8%; Score 103.5; DB 5; Length 379;

Best Local Similarity 24.5%; Pred. No. 2.7;

Matches 38; Conservative 28; Mismatches 44; Indels 45; Gaps 7;

QY 7 IQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLS 59

DB 207 VLKKEASS-----YDYIL-----GWFEFGVPEHKKENMLSHLYVSSKD 247

QY 60 KDTFKIPVKIEEKEENKPTFDVSKKNDPQVNHSQLN-----ESHRKE 106

DB 248 KENISKENDVLDE-KEEAEETEEBELEKNEETESISEBEEEBEEREEKENDKKK 306

QY 107 DLQREHSQKSDTKDVTATVLDKNNISKSTTN 141

DB 307 EOEKESNNENQKDMEA-----QNLISKNNNN 336

RESULT 21

Q9LH98 PRELIMINARY; PRT; 2081 AA.

AC Q9LH98;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Genomic DNA, thaliana chromosome 3, BAC clone: T19N8.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Columbia;

RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Columbia;

RX MEDLINE=20363099; PubMed=10907853;

RA Nakamura Y.;

RT "Structural analysis of Arabidopsis thaliana chromosome 3. II. Sequence features of the regions of 4,251,695 bp covered by ninety P1, TAC and BAC clones.";

RT DNA Res. 7:217-221(2000).

DR EMBL; AP002057; BAB03174.1; -

SQ SEQUENCE 2081 AA; 232851 MW; D3603E1F85EFPF29 CRC64;

Query Match 13.7%; Score 102.5; DB 10; Length 2081;

Best Local Similarity 27.4%; Pred. No. 20;

Matches 43; Conservative 24; Mismatches 63; Indels 27; Gaps 6;

QY 4 TVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGD-----AG 53

DB 1659 TVEINGGEELSTEGSKD-----GKIEGK--EGKENSTKEGSKDDKIEEGMEGKEN 1708

QY 54 YVINLSKDTFKIPVKIEEKEENKPTFDVSK-KKNDPQVNHSQLNSHRKEDLQ- 109

DB 1709 STKESSKDGKINEIHGDKKATMEEGSKDGGTNGTCKSKDSKSVINGVKDDSLKDDSKN 1768

QY 110 ---REHSQKSDTKDVTATVLDKNNISKSTTNPN 143

DB 1769 GDINEINGKEDSVKDNVTETIQGNDNSLTNSTSEPN 1805

RESULT 22

Q9V7J0 PRELIMINARY; PRT; 382 AA.

AC Q9V7J0; Q9GQ81;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE G8421 protein (Aspartyl beta-hydroxylase variant 2).

OS ASPH OR CG8421 OR CG18658.

OC Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkely;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abril J.F., Agbayani A., An H.-J., Andrews-Ffannkuch C., Baldwin D., Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brotter P., Borkova D., Botchan M.R., Bouck J., Brokstein P., Bolshakov S., Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarrey C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nussekern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Fieriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RA "Sequencing of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RA "Annotation of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20564328; PubMed=10956665;  
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
RA Friedman P.A.;  
RT Aspartyl beta -Hydroxylase (Asph) and an Evolutionarily Conserved  
RT Isoform of Asph Missing the Catalytic Domain Share Exons with  
RT Junction.";  
RL J. Biol. Chem. 275:39543-39554(2000).  
DR EMBL; AF289494; AAF58063.2; -;  
DR EMBL; AF289494; AAF58063.2; -;  
DR FlyBase; FBgn0034075; Asph.  
SQ SEQUENCE 382 AA; 43287 MW; 605E3C03ABFC6E8B CRC64;

Query Match 13.6%; Score 101.5; DB 5; Length 382;  
Best Local Similarity 24.5%; Pred. No. 3.8;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
OY 21 EDFILPVYKGELEKGYQFDGW-----EISFGKKGADGYVI-----NLKOTFIK 65  
DB 78 ELDLPTLSRFSK-VFDGWVDEHRDEHGDHVDQPSGEALDDHDEHDDHDEDEE 135

OY 66 PVFKIKBEKEENKPT-----FDVSKKKDNQVNHSQLNESHKEDLOREHSQKSDS 119  
DB 136 PLTELEEELEEREPEEPTDEPADEVEEDEDENNA--GENITAEADAEEEDND 193  
OY 120 TKQVATVLDKNKISSKST 138  
DB 194 EGTVEATVEATTEAT 212  
RESULT 23  
QSV719 ID Q9V719 PRELIMINARY; PRT; 556 AA.  
AC Q9V719; AC Q9V719;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE CG421 protein.  
GN ASPH OR CG421 OR CG18658.  
OS *Drosophila melanogaster* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_taxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarrey C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nussekern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Fieriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RA "Sequencing of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RA "Annotation of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20564328; PubMed=10956665;  
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
RA Friedman P.A.;  
RT Aspartyl beta -Hydroxylase (Asph) and an Evolutionarily Conserved  
RT Isoform of Asph Missing the Catalytic Domain Share Exons with  
RT Junction.";  
RL J. Biol. Chem. 275:39543-39554(2000).  
DR EMBL; AF289494; AAF58063.2; -;  
DR EMBL; AF289494; AAF58063.2; -;  
DR FlyBase; FBgn0034075; Asph.  
SQ SEQUENCE 382 AA; 43287 MW; 605E3C03ABFC6E8B CRC64;

```
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN
RP
SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnikier S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN
RP
SEQUENCE FROM N.A.
RA Adams M.D., Celnikier S.E., Gibbes R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN
RP
SEQUENCE FROM N.A.
RA FlyBase;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS003408; AAF58064.2; -
DR FlyBase; FBgn0034075; Asph.
SQ SEQUENCE 556 AA; 63144 MW; B420980CB6C357A CRC64;

Query Match 13.6%; Score 101.5; DB 5; Length 556;
Best Local Similarity 24.5%; Pred. No. 5;8;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 21 EDPLPVYKGELEKGYQFGW-----EISGFEKKDAGVI-----NLKDTFTK 65
Db 78 ELDLTPLESRSFSK--VFDCGWDEHDEHDGHDVQPSGEALDDHDEHDHDEDEDE 135
QY 66 PVFKTLEEKKEENKPT-----FDVSKKKNQVNHSQLNSHRKEDLQREHSQKSPS 119
Db 136 PLTEELTEELSEEBEETDEPADEBEYDEDEENNA--GENTITAEAESEEBEDND 193

QY 120 TKDVTATVLDDKNISSKST 138
Db 194 EGVTEATVTEATTTEAT 212

RESULT 24
Q9G082 PRELIMINARY; PRT; 785 AA.
AC
Q9G082
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA).
GN ASPH OR CG8421 OR CG18658.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN
RP
SEQUENCE FROM N.A.
RX MEDLINE=20564328; PubMed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved
RT Isoform of Asph Missing the Catalytic Domain Share Exons with
RT Junction.";
```

```
RL J. Biol. Chem. 275:39543-39554 (2000).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Worman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Acabavi A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu R.E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hardek N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidel-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN
RP SEQUENCE FROM N.A.
RA Celnikier S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Ananides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Swirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celnikier S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
```

RESULT 26		
ID	Q9PPL5	PRELIMINARY; PRT; 312 AA.
AC	OC9PPL5	
DT	DT 01-OCT-2000 (TReMBLrel. 15, Created)	
DT	DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)	
DT	DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)	
DE	DE Putative membrane protein.	
GN	GN C70692C.	
OS	OS Campylobacter jejuni.	
OC	OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;	
OC	OC Campylobacteraceae; Campylobacter.	
OX	OX NCBI_TaxId=197;	
RP	RP [1]	
RP	RP SEQUENCE FROM N.A.	
RC	RC STRAIN=NCTC 11168;	
RC	RC MEDLINE=20150912; PubMed=10688204;	
RA	RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,	
RA	RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,	
RA	RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,	
RA	RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,	
RA	RA Whitehead S., Barrrell B.G.;	
RT	RT "The genome sequence of the food-borne pathogen Campylobacter jejuni	
RT	RT reveals hypervariable sequences.";	
RL	RL Nature 403:665-668(2000).	
DR	DR EMBL; AL1139076; CAB72966.1; -.	
KW	KW Complete proteome.	
SQL	SQL SEQUENCE 312 AA; 37221 MW; 0004FA7836A741E8 CRC64;	
Query Match 13.4%; Score 100; DB 16; Length 312;		
Best Local Similarity 25.3%; Pred. No. 4;		
Matches 41; Conservative 32; Mismatches 61; Indels 28; Gap		
QY	7	I QNGKMSSTI---VSEEDFILPVVK-----GELEKGYQPDGWEISGFEGKQDAG
DB	24	INQGPLDDLDREISSDIIURRRFKKTPNKFLEELDEEYESKTKKSNIVLYKED--
QY	56	INL----SKDTFKPVFKIEKKEENKPT---FDVSKKKDNP----QVNHQSQNHSH
DB	81	INVLEEKQSLAKKIFSKWKKRKEENKTKNFKLSRKXANEIKNIQTKTQIQTKNSN
QY	106	EDLQREHSQKSDTKDV--TATVLDKNNISSK--STNNPN 143
DB	141	TTQTQKQKELTNSIEKIQTKTETRIQRPLEIKKLDVKNQPN 182
RESULT 27		
ID	Q81JZ4	PRELIMINARY; PRT; 1130 AA.
AC	OC81JZ4	
DT	DT 01-WAR-2003 (TReMBLrel. 23, Created)	
DT	DT 01-WAR-2003 (TReMBLrel. 23, Last sequence update)	
DT	DT 01-WAR-2003 (TReMBLrel. 23, Last annotation update)	
DE	DE Hypothetical protein.	
GN	GN PF10_0046.	
OS	OS Plasmodium falciparum (isolate 3D7).	
OC	OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	OX NCBI_TaxId=36329;	
RP	RP [1]	
RP	RP SEQUENCE FROM N.A.	
RC	RC STRAIN=3D7;	
RC	RC MEDLINE=22255705; PubMed=12368864;	
RA	RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,	
RA	RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James	
RA	RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,	
RA	RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.	
RA	RA Perteau M., Allen J., Selengut J.J., Haft D., Mather M.W., Vaidya A.B.	
RA	RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.	
RA	RA McFadden G.I., Cummings I.M., Subramanian G.M., Mungall C.	
RA	RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,	
RA	RA Fraser C.M., Barrrell B.;	
RT	RT "Genome sequence of the human malaria parasite Plasmodium	
RT	RT falciparum.";	







DB 194 EGTVEATVEATTEATT 212

RESULT 34

Q815T1 PRELIMINARY; PRT; 558 AA.

ID Q815T1

AC Q815T1;

DT 01-MAR-2003 (TREMBlurel. 23, Created)

DT 01-MAR-2003 (TREMBlurel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlurel. 23, Last annotation update)

DE Hypothetical protein.

GN PFL0600W.

OS Plasmodium falciparum (isolate 3D7).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxId=36329;

[1]

RN SEQUENCE FROM N.A.

RP STRAIN=3D7;

RC MEDLINE=2255705; PubMed=12368864;

RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,

RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,

RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,

RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,

RA Peta M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,

RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,

RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,

RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

RA Fraser C.M., Barrell B.,

RT "Genome sequence of the human malaria parasite Plasmodium

RT falciparum."

RL Nature 419:498-511(2002).

[illegible]

DE Hypothetical 109.0 kDa protein.  
GN T28D5.30 OR AT4G08340.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A.,  
RA Barrall B.G., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K.,  
RA Schueller C.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]

```
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL109819; CAB52556.1; --
DR EMBL; AL161511; CAB77959.1; --
DR InterPro; IPR001760; Opsin.
DR Pfam; PF02902; Peptidase_C48; 1.
DR PROSITE; PS00238; OPSIN; 1.
DR PROSITE; PS0600; ULP_PROTEASE; 1.
KW Hypothetical protein.
SQ SEQUENCE 988 AA; 108955 MW; 01C518587D460EAD CRC64;

Query Match 13.0%; Score 97; DB 10; Length 988;
Best Local Similarity 22.1%; Pred. No. 24;
Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;

QY 9 NGKEMSSSTIVSEEDFILPVYKGELEKGY-----QFDGWEISGFEGKDGAGVYINLSKDTF 63
DB 383 NGRQNSNVQSSVDLSITTYDKVSGVLNVSERDVELVEDDVSAGGLSPVQRDN- 441
QY 64 IKPVFKKIEEKE-----EENKPTFDVSKKDNPOVNHSQLNESHRKEDLQREE 112
DB 442 VEPVGDVRRSGDMPNPSAANNVREGPATFDIMESDNPGRDNVAPWEDHIRSRVQLSP 501
QY 113 HSQKSDSTKDV--ATVLDKNNISSKSTTNNPNK 144
DB 502 HVL-----GAKDVTDSVSDPTDKVGVDVTDASDPT 532

RESULT 36
QY SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
DR EMBL; AL929351; CAD51442.1; --
KW Hypothetical protein.
SQ SEQUENCE 531 AA; 62575 MW; 3A77BD8B6BE4A41 CRC64;

Query Match 12.8%; Score 96; DB 5; Length 531;
Best Local Similarity 27.7%; Pred. No. 14;
Matches 44; Conservative 27; Mismatches 64; Indels 24; Gaps 8;

QY 5 VTIONGKEMSSSTIVSEEDFI-LPVYKGELEKGYFDGWEI-----SGFEGKDGAGVYI 56
DB 17 IEIENKQLIDLKKKBELLSPLEKQEL-----YLLDSPQILLKDNVYKLECKIQNGSMI 73
QY 57 NLSKDTFIKPVFKIE--EKKEEN-KPTFDVSKKDNPOVNHSQLNE--SHRKEDL--- 108
DB 74 QLKSD--IKPSTKKKEENKKNENIKNTDENNKETYDKDVHKNQMDSLCQKKEIVNN 131
QY 109 ---QREEHSQKSDSTKDVTDVLDKNNISSKSTTNNPNK 144
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 30.4847 Seconds  
(without alignments)  
697.707 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSSITVSEDFILPVYK.....ATVLDKNNISKSTNNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03.\*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	696	100.0	773	22 AAB48343	S. pneumoniae Sp13
2	696	100.0	2120	21 AAY81710	Streptococcus pneu
3	696	100.0	2140	24 ABU01020	S. pneumoniae type
4	615	88.4	117	19 AAW55096	Streptococcus pneu
5	615	88.4	117	23 ABP54590	S. pneumoniae Sp04
6	101.5	14.6	564	22 ABB61977	Drosophila melanog
7	98	14.1	665	21 AAB18278	Plasmodium falcipa
8	92.5	13.3	2519	22 ABG15636	Novel human diagno
9	91	13.1	484	21 AAG47777	Arabidopsis thalia

10	89.5	12.9	511	20	AAV35091	Chlamydia pneumonia
11	88.5	12.7	1141	22	AAG85008	Shrimp white spot
12	88	12.6	225	23	ABP73992	Candida albicans e
13	88	12.6	258	22	AAB94584	Human protein sequ
14	88	12.6	746	22	AAG81779	S. epidermidis ope
15	88	12.6	778	23	ABP39023	Staphylococcus epi
16	88	12.6	817	22	AAW79318	Human protein SEQ
17	88	12.6	817	22	AAW79319	Human protein SEQ
18	87	12.5	281	22	ABB64828	Drosophila melanog
19	87	12.5	607	22	AAW39950	Human polypeptide
20	87	12.5	635	21	AA18176	Plasmodium falcipa
21	86.5	12.4	2485	21	AA18176	Plasmodium falcipa
22	86	12.4	1980	21	AA23563	Human sodium chann
23	86	12.4	1980	23	AAQ14927	Human sodium chann
24	85.5	12.3	884	23	ABP73375	Candida albicans e
25	85	12.2	329	21	AAG08296	Arabidopsis thalia
26	85	12.2	339	21	AAG08295	Arabidopsis thalia
27	85	12.2	358	21	AAG08294	Arabidopsis thalia
28	85	12.2	465	23	ABP39011	Staphylococcus epi
29	85	12.2	472	22	AAG81782	S. epidermidis ope
30	85	12.2	720	22	AAG82482	S. epidermidis ope
31	85	12.2	728	23	ABP40123	Staphylococcus epi
32	85	12.2	1166	21	AA18288	Plasmodium falcipa
33	85	12.2	1978	19	AAW69361	Tetradotoxin-sensi
34	85	12.2	1988	19	AAW69362	Tetradotoxin-sensi
35	84.5	12.1	157	23	ABP29728	Streptococcus poly
36	84.5	12.1	161	23	ABP26418	Streptococcus poly
37	84.5	12.1	309	21	AAG35428	Zea mays protein f
38	84.5	12.1	361	21	AAG35427	Zea mays protein f
39	84.5	12.1	390	21	AA335803	Protein involved i
40	84.5	12.1	416	21	AAG35426	Zea mays protein f
41	84.5	12.1	434	23	ABP73309	Candida albicans e
42	84.5	12.1	645	24	ABJ19106	Pathogen specific
43	84.5	12.1	654	24	ABP56879	Staphylococcus epi
44	84.5	12.1	2663	22	AB998612	Human tumour suppr
45	84	12.1	215	23	ABJ11001	Yeast selected int

#### ALIGNMENTS

##### RESULT 1

AAAB48343  
ID AAB48343 standard; Protein; 773 AA.

XX AAB48343;

AC AAB48343;

XX 20-APR-2001 (first entry)

XX S. pneumoniae Sp130 polypeptide.

XX Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;

XX bronchial; lung; blood; infection; immune response; immunotherapy;

XX antibacterial; auditory; vaccine.

XX Streptococcus pneumoniae.

XX WO2000076540-A2.

XX 21-DEC-2000.

XX 09-JUN-2000; 2000WO-US15925.

XX 10-JUN-1999; 99US-0138453.

XX (MEDI-) MED IMMUNE INC.

XX Adamou JE, Choi GH;

XX WPI; 2001-112197/12.

XX N-PSDB; AAC84742.

XX New vaccines comprising Sp128 or Sp130 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by  
 PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
 XX blood infections  
 PS Claim 8; Page 51-54; 54pp; English.  
 XX The invention relates to novel immunogenic polypeptides, Sp128 and Sp130  
 CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
 CC the treatment and prevention of pneumococcal infections, particularly  
 CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
 CC bronchial, lung or blood infections. The antigens are used as immunogenic  
 CC agents to stimulate an immune response. The antisera and antibodies may  
 CC also be used in diagnosing and treating pneumococcal infections.  
 CC Recombinant polypeptides serve as a mechanism for stimulating production  
 CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
 CC as reagents in other processes such as affinity chromatography. The  
 CC present sequence represents the S. pneumoniae Sp130 polypeptide.  
 XX  
 XX Sequence 773 AA;

Query Match 100.0%; Score 696; DB 22; Length 773;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-63;  
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 60  
 DB 640 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 699  
 QY 61 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 120  
 DB 700 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 759  
 QY 121 NNISKSTNNPNK 134  
 DB 760 NNISKSTNNPNK 773

RESULT 2  
 AAY81710  
 ID AAY81710 standard; Protein; 2120 AA.  
 AC AAY81710;

02-JUN-2000 (first entry)

Streptococcus pneumoniae protein sequence ID3.

Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
 bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
 kidney disease; diabetes; immunosuppressive disorder; otitis media;  
 pneumococcal septicaemia; sinusitis; meningitis; therapy.

Streptococcus pneumoniae.

WO200006738-A2.

10-FEB-2000.

27-JUL-1999; 99WO-GB02452.

27-JUL-1998; 98GB-0016336.

19-MAR-1999; 99US-0125329.

(MICK-) MICROBIAL TECHNIQS LTD.

Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;

WPI; 2000-195301/17.

N-PSDB; AA291806.

Streptococcal proteins and polynucleotides useful for diagnosis,  
 treatment and prophylaxis of bacterial infections

PS Claim 2; Page 41-42; 76pp; English.

XX This sequence represents a Streptococcus pneumoniae protein of the  
 CC invention. The proteins (or their homologues, derivatives and/or  
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
 CC compositions comprising the proteins are useful as vaccines and also in  
 CC diagnostic assays. The sequences are useful for the detection or  
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
 CC with them. Agents capable of antagonising, inhibiting or interfering with  
 CC the function or expression of the protein or polypeptide are useful in  
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
 CC infection. As the sequences can be used to treat S. pneumoniae infection,  
 CC they can be used to treat bacterial pneumonia, which has high rates in  
 CC young children, the elderly, and in patients with predisposing conditions  
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
 CC or with immunosuppressive disorders, especially AIDS. They can also be  
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
 CC meningitis.

XX Sequence 2120 AA;

Query Match 100.0%; Score 696; DB 21; Length 2120;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-62;  
 Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 60

DB 1953 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK 2012

QY 61 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 120

DB 2013 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 2072

QY 121 NNISKSTNNPNK 134

DB 2073 NNISKSTNNPNK 2086

RESULT 3

ABU01020

ID ABU01020 standard; Protein; 2140 AA.

AC ABU01020;

11-FEB-2003 (first entry)

S. pneumoniae type 4 strain protein from coding region #590.

Bacterial meningitis; pneumonia; sepsis; otitis media;  
 ear infection; antiinflammatory; antibacterial; immunostimulant;  
 auditory; respiratory; gene therapy; vaccine.

Streptococcus pneumoniae type 4 strain.

WO200277021-A2.

03-OCT-2002.

27-MAR-2002; 2002WO-IB02163.

27-MAR-2001; 2001GB-0007658.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Masignani V, Tettelin H, Fraser C;

WPI; 2003-040579/03.

N-PSDB; ABX06302.

New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 useful as medicaments for treating or preventing a disease or infection  
 due to streptococcus bacteria, such as pneumonia, sepsis, otitis media

```

PT or ear infection -
XX Claim 1; SEQ ID No 1180; 56pp; English.
XX
XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,
XX expressed from 2469 of 2489 identified DNA coding regions from the
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX AB556454. Also included are an antibody which binds one of the
XX proteins, treating a patient by administering the protein, DNA or
XX antibody (in a composition), a kit comprising first and second primers,
XX which are the nucleic acid cited above or fragments between nucleotides
XX 8-100 of a sequence not defined in the specification, for amplifying a
XX target sequence contained within a Streptococcus nucleic acid sequence,
XX where the first primer is substantially complementary to the target
XX sequence and the second primer is substantially complementary to the
XX complement of the target sequence, and where the parts of the primers
XX having substantial complementarity define the termini of the target
XX sequence to be amplified, assay comprising contacting a test compound
XX with the protein, and determining whether the test compound binds to the
XX protein and a Streptococcus pneumoniae bacterium, where one or more
XX genes encoding the proteins has been rendered inactive. The proteins,
XX nucleic acid molecules, antibody and compositions are useful as
XX medicaments for treating or preventing a disease or infection due to
XX streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
XX sepsis, otitis media or ear infection. They are also useful in developing
XX vaccines, diagnostics and antibiotics. The methods are useful for
XX identifying immunodominant proteins. The present sequence is one of
XX the 2469 proteins expressed by the identified coding regions from the
XX genomic sequence.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2140 AA;
Query Match 100.0%; Score 696; DB 24; Length 2140;
Best Local Similarity 100.0%; Pred. No. 1.3e-62; Mismatches 0; Indels 0; Gaps 0;
Matches 134; Conservative 0;
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKK 60
DB 1973 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKK 2032
QY 61 IEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 120
DB 2033 IEKKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDK 2092
QY 121 NNISKSTNNPNK 134
DB 2093 NNISKSTNNPNK 2106
RESULT 4
AAW55096
ID AAW55096 standard; Protein; 117 AA.
XX
XX AAW55096;
XX
XX 02-OCT-1998 (first entry)
XX Streptococcus pneumoniae SP0043 protein.
XX
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX detection; pneumonia; otitis media; meningitis.
XX
XX Streptococcus pneumoniae.
XX
XX WO9818930-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US19422.
XX
XX 31-OCT-1996; 96US-0029960.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX WPI; 1998-272224/24.
XX N-PSDB; AAV27357.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus
XX pneumoniae - or their epitope-containing fragments, useful in
XX protective or therapeutic vaccines, and for diagnosis
XX
XX Claim 11; Page 62; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX can be useful in vaccines for inducing protective antibodies against
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX are used to detect Streptococcus infection (by usual hybridisation or
XX amplification methods), also for isolating Streptococcus genes or their
XX allelic variants. The protein can be used similarly to detect specific
XX antibodies in standard immunoassays, especially for diagnosing or
XX monitoring infections. Antibodies which bind the protein are used to
XX detect corresponding antigens, to purify the protein and for passive
XX immunisation (optionally coupled to a toxin). Vaccines are administered,
XX e.g. by injection, orally or through the skin, typically at 0.01-1000
XX (especially 10-300) mu g/ml per dose.
XX
XX Sequence 117 AA;
Query Match 88.4%; Score 615; DB 19; Length 117;
Best Local Similarity 100.0%; Pred. No. 6.1e-56;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 18 YKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 77
DB 1 YKGELEKGYQFDGWEISGFEKGKDGAGYVINLSKDTFIKPVFKKIEKKKEENKPTFDVSK 60
QY 78 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 134
DB 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117
RESULT 5
*ABP54590
ID ABP54590 standard; Protein; 117 AA.
XX
XX ABP54590;
XX
XX 04-SEP-2002 (first entry)
XX
XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.
XX
XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX antibacterial; Streptococcal infection; detection.
XX
XX Streptococcus pneumoniae.
XX
XX US2002061545-A1.
XX
XX 23-MAY-2002.
XX
XX 22-JAN-2001; 2001US-0765272.
XX
XX 30-OCT-1997; 97US-0961083.
XX (CHOI/) CHOI G H.
XX (KUNS/) KUNSCH C A.
XX

```





XX Disclosure; Page 321-322; 577pp; English.

PS The present invention describes proteins and their fragments (I) encoded

XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.

CC Also described are: (i) nucleotide sequences (II) encoding (I); and (2)

CC vaccines against P. falciparum infection comprising (I) or (II).

CC (I) and (II) are useful for the development of vaccines against

CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal

CC antibody raised to immunogens comprising the sequences of (I), are

CC useful in the detection of infection with P. falciparum. Furthermore,

CC (II) especially when they are rifins or secreted or membrane proteins)

CC can aid the identification of drugs to treat or prevent P. falciparum

CC infection, or they can be used to identify drug resistance in

CC P. falciparum. Sequencing of the plasmodium chromosome 2 and the

CC subsequent identification of proteins encoded by it will help to expand

CC our understanding of parasite biology, a process hampered by the

CC complexity of the parasitic lifecycle, and provide new targets for

CC vaccine and drug development. Parasite resistance to drugs and mosquito

CC resistance to insecticides have led to a resurgence of malaria in many

CC parts of the world, and there is a pressing need for vaccines and new

CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide

CC and protein sequences given in the present invention, but which are not

CC specifically mentioned within the specification.

XX Sequence 665 AA;

Query Match 14.1%; Score 98; DB 21; Length 665;

Best Local Similarity 27.1%; Pred. No. 0.26; Mismatches 29; Indels 42; Gaps 7;

Matches 35; Conservative 23;

QY 23 EKGQFDGWEI--SGFEGKGDAGYVNLKDTFVFKKIEKKE-----EENK 70

DB 158 EKGQ-----DISNNAENKQD-----VREGVKELEKKKEKISDDHKVEENK 201

QY 71 PTFD-----VSKKDNPNVHNSQLNESHKEDLQR-EHHSQKSDSTKDVATVLDKNNISS 125

DB 202 KSDDHKVEENKSDDHKVEENKSDDHKIEEVKVEEHEEDDEE-----DKKEKKS 252

QY 126 KSTTNPNK 134

DB 253 ENKNKDNK 261

RESULT 8

ABG16636

ID ABG16636 standard; Protein; 2519 AA.

AC ABG16636;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #16627.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

PD 30-MAR-2001; 2001WO-US08631.

PF 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

XX

DR N-PSDB; AAS80823.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

XX Claim 20; SEQ ID NO 46995; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2519 AA;

Query Match 13.3%; Score 92.5; DB 22; Length 2519;

Best Local Similarity 31.0%; Pred. No. 5.9; Mismatches 19; Indels 13; Gaps 3;

Matches 22; Conservative 19;

QY 48 LSKDTFIKVPFKTEEEKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLQRE-----E 102

DB 689 VKKETVKVP-----EDKKEEKPKKEVAKEDKTPF---KKEKPKKEVKVEKKEIK 740

QY 103 HSQKSDSTKDV 113

DB 741 KEKKEPKKEV 751

RESULT 9

AAG47777

ID AAG47777 standard; Protein; 484 AA.

AC AAG47777;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 60255.

DE Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139751.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142877.  
PR 13-JUL-1999; 99US-0143354.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.

```
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160983.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161362.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 13.1%; Score 91; DB 21; Length 484;
Best Local Similarity 19.3%; Pred. No. 0.89;
Matches 29; Conservative 29; Mismatches 59; Indels 34; Gaps 4;

QY 18 YKGELEKGYQDFWEISGFE-----GKIDAGVIVNLKDTIKPVFKIEKKEEN 69
DB 114 YVQDLARRIRYDE-EATGSAQRIDHPNKNVIGITEKAFENSPTEETSHRVDDNKRINN 172

QY 70 KPTFDVSKKDN-----PQVNHSLN-----SHRKEDLQREHS 104
DB 173 QKNFATAKSSNAVSRVSGADHKGAEVNGKPMENRQVQTESAESKSRKENVTKSEK 232

QY 105 QKSDTKDVTATVLDKNNISSKSTNNPNK 134
DB 233 RDQGVKYTEAKDKDRNKEKKEKTESINK 262

RESULT 10
AAV35091
ID AAV35091 standard; Protein; 511 AA.
AC AAV35091;
XX
XX
DT 13-SEP-1999 (first entry)
DE
DE Chlamydia pneumoniae transmembrane protein sequence.
XX
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
PN WO9927105-A2.
XX
XX 03-JUN-1999.
XX
XX 20-NOV-1998; 98WO-IB01890.
XX
XX 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
XX (GEST ) GENSET.
PA
```

```
XX Griffais R;
XX WPI; 1999-357842/30.
XX
XX Genome sequence of Chlamydia pneumoniae
XX
XX Page 975-976; Disclosure; 1912pp; English.
XX
XX AAY34584-Y35879 represent the proteins encoded by all the open reading
frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
SQ Sequence 511 AA;

Query Match 12.9%; Score 89.5; DB 20; Length 511;
Best Local Similarity 24.5%; Pred. No. 1.4;
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

QY 54 IKPVFKKIEKKEENKPTFD-----VSKKDNQVNHSLNESHK 95
DB 95 VKGVFKTTPQARPEVSSPRLPSHVQHGQRLPGLGLEGFRDQKRSENPADLGKMKRSYSD 154

QY 96 EDLQREHSQKSDTKDVTATVLDKNNISSKSTT 129
DB 155 GDLRVGHDSNEDSTEDSRS---EGGEPSSKSS 185

RESULT 11
AAG85008
ID AAG85008 standard; Protein; 1141 AA.
XX
AC AAG85008;
XX
DT 11-SEP-2001 (first entry)
DE
DE Shrimp white spot Bacilliform virus (WSBV) protein 99.
XX
KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW antiviral agent; gene expression; antisense construct;
KW transgenic viral resistant shrimp.
XX
OS White spot syndrome virus.
XX
PN WO200138351-A2.
XX
XX 31-MAY-2001.
XX
XX 08-NOV-2000; 2000WO-US28888.
XX
XX 24-NOV-1999; 99CN-0124717.
XX
XX (PENY-) PE CORP NY.
XX (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
XX (SINO-) SINOGENOMAX CO LTD.
XX
XX Xu X, Yang F, He J, Pham L, He M., Ye Y, Shen Y, Kodira C;
XX WPI; 2001-355877/37.
XX N-PSDB; AAH62788.
XX
XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus
XX (WSBV), useful for producing viral polypeptides that can be used to
XX screen for agents that are useful for treating WSBV infection -
XX
```





```

PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2002-381255/41.
DR N-PSDB; ABN91568.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure; SEQ ID 3868; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
XX Sequence 778 AA;
SQ
Query Match 12.6%; Score 88; DB 23; Length 778;
Best Local Similarity 25.7%; Pred. No. 3.5;
Matches 37; Conservative 21; Mismatches 50; Indels 36; Gaps 7;
QY 4 SSTIASEDFILPVYKGELEKGYQPDG---EISGPE-----GKKDAGYVIN--LSKD 51
DB 643 SGTKVLPHSKVLMITDGLTMP-DMTGWTKEVDLAFEDLTJKVSTKGNFVNTQSISKG 701
QY 52 TRIKPVFKKIEBKKEBENKPTFDVS-----KXKDNPOVNHSQLNESHRKEDLQREHSQKS 107
DB 702 QIHK-----NKKIEVLSLSAEDTDDQEKTDSDDNKSKKDADEHSNNTS 748
QY 108 DSTKQVATVLDKNNISSKSTNN 131
DB 749 SSTKN-----DKSNADSKNDSDD 766
RESULT 16
AAM79318
ID AAM79318 standard; Protein; 817 AA.
XX
AC AAM79318;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 2964.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
DE WO200157190-A2.
XX
PN WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX
XX 27-APR-2000; 2000US-0560875.
XX
XX 20-JUN-2000; 2000US-0598075.
XX
XX 19-JUL-2000; 2000US-0620325.
XX
XX 01-SEP-2000; 2000US-0654936.
XX
XX 15-SEP-2000; 2000US-0663561.
XX
XX 20-OCT-2000; 2000US-0693325.
XX
XX 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK52451.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 215; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 817 AA;
SQ
Query Match 12.6%; Score 88; DB 22; Length 817;
Best Local Similarity 27.3%; Pred. No. 3.7;
Matches 33; Conservative 21; Mismatches 55; Indels 12; Gaps 4;
QY 20 GELEKGYQFDGWEISG--FEKGDAGYVINLSKDTFIRPVFKKIEKKEENKPTFDVSK 77
DB 196 GOEKKEQSFKSWEASGKHQVSKPAVLEQRKQDTSKLRSTLPEKQKQEKSKSPSPSQ 255
QY 78 -KKDNPOVNHSQLNESHRKED-----LQREHSQKSDSTKQVATVLDKNNISSKST 128
DB 256 WKQDTPKSKAGYVQSEHKQETPKLPVQLQKEQ-DPKKQTPKSWTPSMQSQNTTKSWT 314
QY 129 T 129
DB 315 T 315
RESULT 17
AAM79319
ID AAM79319 standard; Protein; 817 AA.
XX
AC AAM79319;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 2965.
XX
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
DE WO200157190-A2.
XX
PN WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
XX
XX 27-APR-2000; 2000US-0560875.

```





```
PF 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR N-PSDB; AAI59106.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 3095; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAW42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 607 AA;
SQ
Query Match 12.5%; Score 87; DB 22; Length 607;
Best Local Similarity 33.3%; Pred. No. 3.2;
Matches 20; Conservative 12; Mismatches 28; Indels 0; Gaps 0;
QY 49 SKDTFTKPKVFKTEKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSD 108
DB 373 SPDNCCNELFKKKKKKEEKKKEEKKRRRRRRKREKREKREKREKREKREK 432
RESULT 20
AAB18176
ID AAB18176 standard; Protein; 635 AA.
XX
XX AAB18176;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:33.
DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
XX Plasmodium falciparum.
OS
XX WO200025728-A2.
PN
XX 11-MAY-2000.
PD
XX 05-NOV-1999; 99WO-US26796.
PF
```

```
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
PI WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 85-86; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAT70078 to AAT70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
XX Sequence 635 AA;
SQ
Query Match 12.5%; Score 87; DB 21; Length 635;
Best Local Similarity 23.8%; Pred. No. 3.4;
Matches 30; Conservative 20; Mismatches 52; Indels 24; Gaps 4;
QY 22 LEKGYQFGWRTSGEGKKDAGYVNLKSKDTFKFVKFKIEKKEENKPTFDVSKKDN 81
DB 202 LQKKNYIODEEDNETIRSDSKLDIYDSQSKDIMSSSPNKEES-----MSSDNHN 256
QY 82 PQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLD-----KNNISKS 127
DB 257 KDINSS---DNQNKDINSDDHNM-NDSTNESTTTLSLTSTNTNRNKKRKNININ 311
QY 128 TTNPN 133
DB 312 NNNNSN 317
RESULT 21
AAB18172
ID AAB18172 standard; Protein; 2485 AA.
XX
XX AAB18172;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related protein SEQ ID NO:29.
DE Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
XX Plasmodium falciparum.
OS
```



XX WPI; 2002-393394/42.  
DR N-PSDB; AAL42751.  
XX Sodium channel alpha subunits SCN12A and SCN8A of human nervous system  
PT together with splicing variants, useful in studying physiological  
PT mechanism relating to excitatory cells and in drug development  
XX  
XX Claim 3; Page 97-110; 118pp; Japanese.  
XX The invention comprises the amino acid and coding sequence of three  
CC sodium channel subunits from the human nervous system (SCN12A, SCN12A-s  
CC and SCN8A). The SCN12A/SCN12A-s gene is found on chromosome 3p23-21.3 and  
CC the SCN8A gene is found on chromosome 12q13.1. The sodium ion channel  
CC subunit proteins are useful in studying physiological mechanisms relating  
CC to excitatory cells and in drug development. The sodium ion channel  
CC subunit proteins are useful for treating diseases such as familial  
CC hyperglycaemia, QT extending syndrome type 3 and motor endplate diseases.  
CC The present amino acid sequence represents the human SCN8A sodium channel  
CC subunit.  
XX  
XX Sequence 1980 AA;  
Query Match 12.4%; Score 86; DB 23; Length 1980;  
Best Local Similarity 26.2%; Pred. No. 20;  
Matches 28; Conservative 23; Mismatches 48; Indels 8; Gaps 4;  
QY 27 QFDGWEISGFEGKQAGYVILNLSKDTFIKPVFKIEEKEENKPTFDVSKKDNPNQVNH 86  
Db 993 EMNNLQISVIRIKGVAVWT-KLVHAFMQAHFK---QREADEVKPLDELYEKKANCIAH 1048  
QY 87 SOLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPN 133  
Db 1049 TGA-DIHRNGDFQKNGGTGSGTGSVEKYIIDEDHM---SPINPN 1091  
RESULT 24  
ABP73375  
ID ABP73375 standard; Protein; 884 AA.  
XX  
AC ABP73375;  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Candida albicans essential protein SEQ ID NO 7212.  
XX  
KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;  
KW signal transduction; DNA replication; cell division; growth;  
KW proliferation; Candida albicans; fungicide; antifungal.  
XX  
OS Candida albicans.  
XX  
PN WO200253728-A2.  
XX  
PD 11-JUL-2002.  
XX  
XX 26-DEC-2001; 2001WO-US9486.  
XX  
PR 29-DEC-2000; 2000US-259128P.  
PR 20-FEB-2001; 2001US-0792024.  
PR 22-AUG-2001; 2001US-314050P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
XX  
XX WPI; 2002-566694/60.  
DR N-PSDB; AB231925.  
XX  
XX Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele  
PT of a gene and placing other allele of the gene under conditional  
PT expression

XX Claim 44; SEQ ID NO 7212; 167pp + Sequence Listing; English.  
XX The invention relates to constructing (M1) a strain of diploid fungal  
CC cells in which both alleles of a gene are modified, comprising modifying  
CC one allele by insertion or replacement by a cassette having an  
CC expressible selectable marker and modifying other allele by  
CC recombination, of a promoter replacement fragment with a heterologous  
CC promoter, so that expression of the second allele is regulated by the  
CC promoter. (M1) is useful for constructing a strain of diploid fungal  
CC cells in which both alleles of a gene are modified. The diploid fungal  
CC cells having both alleles modified are useful for identifying a gene that  
CC is essential to the survival or growth of a fungus, a gene that  
CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
CC that contributes to the resistance of a diploid fungus to an antifungal  
CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
CC and for identifying a therapeutic agent for treatment of a mammalian  
CC disease. (M1) is useful for identifying a compound which modulates the  
CC activity of a gene product, preferably enzymatic activity, carbon  
CC compound catabolism, biosynthetic, transporter, transcriptional,  
CC translational, signal transduction, DNA replication and cell division  
CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office.  
XX  
XX Sequence 884 AA;  
Query Match 12.3%; Score 85.5; DB 23; Length 884;  
Best Local Similarity 25.0%; Pred. No. 7.6;  
Matches 26; Conservative 23; Mismatches 46; Indels 9; Gaps 3;  
QY 34 SGFEGKDGAGYVILNLSKDTFIKPVFKIEEKEENKPTFDVSK--KKDNPNQVNHQ 88  
Db 14 SGQEVKNNINQTSLSKEEKLKREQLWRQKQQQQQQQEQEVQNKAKTSTNTS 73  
QY 89 LNESHRKEDLQREE---HSQKSDTKDVTATVLDKNNISSKST 128  
Db 74 TEEVKKSQRQRIEWEKRLQKQKQATKEKTTTITIKKVHSQTT 117  
RESULT 25  
AAG08296  
ID AAG08296 standard; Protein; 329 AA.  
XX  
AC AAG08296;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 5775.  
DE  
DE Arabidopsis thaliana.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134258.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149358.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155139.  
PR 24-SEP-1999; 99US-0155486.  
PR 28-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158212.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.

```
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match      12.2%; Score 85; DB 21; Length 329;
Best Local Similarity 23.8%; Pred. No. 2.2;
Matches 24; Conservative 22; Mismatches 33; Indels 22; Gaps 4;

QY 16 PVYGELEKGVQDFGWEISGFEKGDAGVINLSKDTIKPVFKIEKKKEENKPTDVF 75
Db 213 PIMERSLKRHVADWDQVD-----INLKVGGQQDV--KLEEKKEEEKQ--DM 257
QY 76 SKKNPNQVNSQLNESHKEDLQREHSQKSDSTKDVAT 116
Db 258 SNEEDEE-----EEEEKQMSDEDDKBEEDQESEK 291

RESULT 26
AAG08295
ID AAG08295 standard; Protein; 339 AA.
XX
AC AAG08295;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 5774.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
```

PR 19-JUL-1999;	99US-0144335.	PR 14-OCT-1999;	99US-0159330.
PR 20-JUL-1999;	99US-0144352.	PR 14-OCT-1999;	99US-0159331.
PR 20-JUL-1999;	99US-0144632.	PR 14-OCT-1999;	99US-0159637.
PR 20-JUL-1999;	99US-0144884.	PR 14-OCT-1999;	99US-0159638.
PR 21-JUL-1999;	99US-0144814.	PR 18-OCT-1999;	99US-0159584.
PR 21-JUL-1999;	99US-0145086.	PR 21-OCT-1999;	99US-0160741.
PR 21-JUL-1999;	99US-0145088.	PR 21-OCT-1999;	99US-0160767.
PR 22-JUL-1999;	99US-0145085.	PR 21-OCT-1999;	99US-0160770.
PR 22-JUL-1999;	99US-0145087.	PR 21-OCT-1999;	99US-0160814.
PR 22-JUL-1999;	99US-0145192.	PR 21-OCT-1999;	99US-0160815.
PR 23-JUL-1999;	99US-0145145.	PR 22-OCT-1999;	99US-0160980.
PR 23-JUL-1999;	99US-0145218.	PR 22-OCT-1999;	99US-0160981.
PR 23-JUL-1999;	99US-0145224.	PR 22-OCT-1999;	99US-0160989.
PR 26-JUL-1999;	99US-0145276.	PR 25-OCT-1999;	99US-0161404.
PR 27-JUL-1999;	99US-0145913.	PR 25-OCT-1999;	99US-0161405.
PR 27-JUL-1999;	99US-0145918.	PR 25-OCT-1999;	99US-0161406.
PR 27-JUL-1999;	99US-0145919.	PR 26-OCT-1999;	99US-0161359.
PR 28-JUL-1999;	99US-0145951.	PR 26-OCT-1999;	99US-0161360.
PR 02-AUG-1999;	99US-0146386.	PR 26-OCT-1999;	99US-0161361.
PR 02-AUG-1999;	99US-0146388.	PR 28-OCT-1999;	99US-0161920.
PR 02-AUG-1999;	99US-0146389.	PR 28-OCT-1999;	99US-0161992.
PR 03-AUG-1999;	99US-0147038.	PR 28-OCT-1999;	99US-0161993.
PR 04-AUG-1999;	99US-0147204.	PR 29-OCT-1999;	99US-0162142.
PR 04-AUG-1999;	99US-0147302.		
PR 05-AUG-1999;	99US-0147192.		
PR 05-AUG-1999;	99US-0147260.		
PR 06-AUG-1999;	99US-0147303.		
PR 06-AUG-1999;	99US-0147416.		
PR 09-AUG-1999;	99US-0147493.		
PR 09-AUG-1999;	99US-0147935.		
PR 10-AUG-1999;	99US-0148171.		
PR 11-AUG-1999;	99US-0148319.		
PR 12-AUG-1999;	99US-0148341.		
PR 13-AUG-1999;	99US-0148563.		
PR 13-AUG-1999;	99US-0148684.		
PR 16-AUG-1999;	99US-0149368.		
PR 17-AUG-1999;	99US-0149175.		
PR 18-AUG-1999;	99US-0149426.		
PR 20-AUG-1999;	99US-0149723.		
PR 20-AUG-1999;	99US-0149723.		
PR 20-AUG-1999;	99US-0149929.		
PR 23-AUG-1999;	99US-0149907.		
PR 23-AUG-1999;	99US-0149930.		
PR 25-AUG-1999;	99US-0150566.		
PR 26-AUG-1999;	99US-0150884.		
PR 27-AUG-1999;	99US-0151065.		
PR 27-AUG-1999;	99US-0151066.		
PR 27-AUG-1999;	99US-0151080.		
PR 30-AUG-1999;	99US-0151303.		
PR 31-AUG-1999;	99US-0151438.		
PR 01-SEP-1999;	99US-0151930.		
PR 07-SEP-1999;	99US-0152363.		
PR 10-SEP-1999;	99US-0153070.		
PR 13-SEP-1999;	99US-0153758.		
PR 15-SEP-1999;	99US-0154018.		
PR 16-SEP-1999;	99US-0154039.		
PR 20-SEP-1999;	99US-0154779.		
PR 22-SEP-1999;	99US-0155139.		
PR 23-SEP-1999;	99US-0155486.		
PR 24-SEP-1999;	99US-0155659.		
PR 28-SEP-1999;	99US-0156458.		
PR 29-SEP-1999;	99US-0156596.		
PR 04-OCT-1999;	99US-0157117.		
PR 05-OCT-1999;	99US-0157753.		
PR 06-OCT-1999;	99US-0157865.		
PR 07-OCT-1999;	99US-0158029.		
PR 08-OCT-1999;	99US-0158232.		
PR 12-OCT-1999;	99US-0158369.		
PR 13-OCT-1999;	99US-0159293.		
PR 13-OCT-1999;	99US-0159294.		
PR 13-OCT-1999;	99US-0159295.		
PR 14-OCT-1999;	99US-0159329.		
Query Match 12.2%; Score 85; DB 21; Length 339;			
Best Local Similarity 23.8%; Pred. No. 2.3;			
Matches 24; Conservative 22; Mismatches 33; Indels 22; Gaps 4;			
Qy 16	PVYKGELEKGYQFGWEISGPEGKKDAGYVNLKDTFKVFKIEBKKEENKPTDQV 75		
Db 223	PIMERSLKRHVEADDQVD-----INLKYKGOQDV--KLEEKKEEKQ--DM 267		
Qy 76	SKKDNQVNHSQLNESHKEDLQREHSQKSDSTKQVAT 116		
Db 268	SNEDEEE-----EEEKQDMSEEDDKBEEDQEEBEKT 301		
RESULT 27			
ID	AAG08294		
XX	AAG08294 standard; Protein; 358 AA.		
AC	AAG08294;		
XX			
DT	17-OCT-2000 (first entry)		
XX			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 5773.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999;		
PR	05-MAR-1999;		
PR	09-MAR-1999;		
PR	23-MAR-1999;		
PR	25-MAR-1999;		
PR	25-MAR-1999;		
PR	01-APR-1999;		
PR	06-APR-1999;		
PR	08-APR-1999;		
PR	16-APR-1999;		
PR	19-APR-1999;		
PR	21-APR-1999;		
PR	23-APR-1999;		



PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0137222.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
Query Match 12.2%; Score 85; DB 21; Length 358;  
Best Local Similarity 23.8%; Pred. No. 2.5;  
Matches 24; Conservative 22; Mismatches 33; Indels 22; Gaps 4;  
QY 16 PVYGELEKGYOFDCWEISGPEGKDGAVINLSKDTFKVPFKIEEKEEENKPTFDV 75  
DB 242 PIMERSLKRHVEADWDQVD-----INLKVGGQDVV-KLEEKEEKEEQ--DM 286  
QY 76 SKKDNQPNVHNSQLNSHRKEDLQREHSQKSDSTKDVAT 116  
DB 287 SNEEDEEE-----EEEEKQDMSEDDKEEDEQEEREKT 320  
RESULT 28  
ABP39011  
ID ABP39011 standard; Protein; 465 AA.  
AC ABP39011;  
XX  
XX  
DT 24-JUL-2002 (first entry)  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3856.  
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy.  
XX  
XX Staphylococcus epidermidis.  
XX US6380370-B1.  
XX  
XX  
PD 30-APR-2002.  
XX  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Doucette-Stamm LA, Bush D;  
XX  
XX WPI; 2002-381255/41.  
DR N-PSDB; ABN91556.  
XX  
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 3856; 267pp; English.  
XX  
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences

CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 465 AA;  
Query Match 12.2%; Score 85; DB 23; Length 465;  
Best Local Similarity 32.6%; Pred. No. 3.5;  
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;  
QY 61 IEEKKEEENKPTFDVSKKDNQPNVHNSQLNSHRKEDLQ-----REEHSQKSDS--- 109  
DB 3 MEENKNQPNKE--NNSNKDDNA---THLNSHRNEDLELFRNKNARRRRRIDNOSK 56  
QY 110 TKDVTAT-----VLDDNNISSKSTTNPNK 134  
DB 57 EXDATSTOSQLETKPMDKFDINHKS--HNQNK 86  
RESULT 29  
AAG81782  
ID AAG81782 standard; Protein; 472 AA.  
XX  
AC AAG81782;  
XX  
XX 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis open reading frame protein sequence SEQ ID NO:658.  
XX  
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
KW vaccination; endocarditis.  
XX  
XX Staphylococcus epidermidis.  
XX WO200134809-A2.  
XX  
XX 17-MAY-2001.  
XX  
XX 09-NOV-2000; 2000WO-US30782.  
XX  
XX 09-NOV-1999; 99US-0164258.  
XX  
XX (GLAX) GLAXO GROUP LTD.  
XX  
XX Kimmerly WJ;  
XX  
XX WPI; 2001-316495/33.  
DR N-PSDB; AAH52632.  
XX  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
XX useful for vaccinating against infections, e.g. endocarditis -  
XX  
XX Claim 18; Page 210; 2188pp; English.  
XX  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
XX (I) and (II) can have antibacterial activity and therefore can be used  
XX in vaccination. The nucleic acids (I) may be used to produce the  
XX S. epidermidis polypeptides (II) via the production of vectors  
XX containing them which are used to produce hosts cells which express the  
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
XX used to vaccinate subjects and to raise antibodies against the bacteria.  
XX The polypeptides may also be used to assay for other inhibitors of their  
XX activity and therefore identify compounds that may be used for the  
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA  
XX polynucleotide sequences from the present invention. AAH55091 to

AAH55098 represent oligonucleotide sequences and primers which are used in the present invention.

N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present for SEQ ID NO:4455 to 4464.

```

AA      SQ      Sequence      472 AA;

Query Match          12.2%; Score 85; DB 22; Length 472;
Best Local Similarity 32.6%; Pred. No. 3.6;
Matches 30; Conservative 15; Mismatches 21; Indels 26; Gaps 6;

Qy    61 IEEKKEBENKPTFDVSKKKONPQVNSQLNESHKEDLQ-----REEHSQKSDS--- 109
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db    3 MEENKNQPKE--NWSNKDDNA----THLNDSHRNEDELEFPRNKNARQRERRRIDNQSK 56

Qy    110 TKQVTAT-----VLDKNNISSKSTNNPNK 134

Db    57 EKATSTOSOLETKPMDFKFLDNHKS--HNQNK 86

```

RESULT 30  
AAG82482  
ID AAG82482 standard: Protein: 720 AA.

AC	ARG82482;
XX	
XX	03-SEP-2001 (first entry)
DT	
XX	
XX	
DE	S. epidermidis open reading frame protein sequence SEQ ID NO:2058.
XX	
XX	Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW	vaccination; endocarditis.
KW	
XX	
OS	Staphylococcus epidermidis.
XX	
XX	
PN	WQ200134809-A2.

PD	17-MAY-2001.	
XX		
PF	09-NOV-2000; 2000WO-US30782.	
XX		
XX		
PR	09-NOV-1999; 99US-0164258.	
XX		
XX		
PA	(GLAX ) GLAXO GROUP LTD.	
XX		
PI	Kimmerly WJ;	
XX		
XX		
DR	WPI; 2001-316495/33.	
DR	N-PSDB; AAH5332.	

XX Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX  
PS Claim 18: Page 558: 2188pp: English.

AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*. (I) and (II) can have antibacterial activity and therefore can be used in vaccination. The nucleic acids (I) may be used to produce the *S. epidermidis* polypeptides (II) via the production of vectors containing them which are used to produce hosts cells which express the polypeptides. The polypeptides (II) (and/or nucleic acids) may then be used to vaccinate subjects and to raise antibodies against the bacteria. The polypeptides may also be used to assay for other inhibitors of their activity and therefore identify compounds that may be used for the treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to AAH5090 represent specifically claimed *S. epidermidis* genomic DNA polynucleotide sequences from the present invention. AAH5091 to AAH5098 represent oligonucleotide sequences and primers which are used in the exemplification of the present invention.

N.B. The present invention specifically claims all the polynucleotide sequences given in the sequence listing of the present specification, however the sequence listing only goes up to SEQ ID NO:4454 so even though sequences are given in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present for SEQ ID NO:4455 to 4464.

Sequence	720 AA;
SO	

Query Match: 12.2%; Score 85; DB 22; Length 720;  
Best Local Similarity 28.0%; Pred. No. 6.4;

	Matches	23	Conservative	15	Mismatches	36	Indels	8	Gaps
Qy	61	1	EEKKEEENKPT	PFVSKKCD	-----	NPQVNHSQLMESHKRLD	QREHHSQKSDSTK	-----	112
		:					:		
Db	38	LEEEI	KALDKKFKASAK	TKNTKONT	NNHOKSNKNNKNSND	KDKKKQSNKSKPTKKKQEN	97		

113 VTATVLDKNNISSKSTTNPNK 134 QY

Db 98 NKGKQONKNKTNKNQKNKNK 119

RESULT 31  
ABP40123  
ID ABP40123 standard: Protein: 728 AA.

XX AC ABP40123:

DT 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4968.

KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KW antibacterial; gene therapy.

OS Staphylococcus epidermidis.

PN US6380370-B1.

PD 30-APR-2002.

PF 13-AUG-1998; 98US-0134001.

PR 14-AUG-1997; 97US-055779P.

XX  
XXXX  
XX

XX  
11  
10  
9  
8  
7  
6  
5  
4  
3  
2  
1

DR N-PSDB; ABN92668.

Novel isolated nucleic acid

XX  
XX

2 XX

CC frame (ORF) nucleic acid seq

antibacterial activity and

particularly *S. epidermidis*

CC cycle or inhibit S. epidermidis

specification, but was obtained

XX

1000

Best Local Similarity	28.0%
-----------------------	-------







PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 22-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 12.1%; Score 84.5; DB 21; Length 309;

Best Local Similarity 24.8%; Pred. No. 2.3;  
Matches 32; Conservative 16; Mismatches 46; Indels 35; Gaps 5;

Qy 15 LPVYGELEKGYQFDGWEISGPEGKK-----DAGYVINLS-----KDTFIKPVFKKI 61  
Db 141 LPVYK-----DNDSFTVKRISDLQASKESPARAVASDKGTFVPSSSSQEKDPVAKTILNKV 196  
Qy 62 EEKKEENKPTFDVSKKKNPQVN-----HSQLNESHKKE-----DLOREH 103  
Db 197 KEKSDDEGKPLPAEFDGKENPVANSKNDKSDSGVDRSRERERGRGRDARGRDSRDRSR 256  
Qy 104 SOKSDSTKD 112  
Db 257 GRESDRERD 265

RESULT 38  
AAG35427  
ID AAG35427 standard; Protein; 361 AA.



XX AAG35427; 99US-0139462.  
XX AC 18-JUN-1999; 99US-0139463.  
XX DT 18-JUN-1999; 99US-0139750.  
XX XX 18-JUN-1999; 99US-0139763.  
XX DE 21-JUN-1999; 99US-0139817.  
XX XX 22-JUN-1999; 99US-0139899.  
XX KW 23-JUN-1999; 99US-0140353.  
XX KW 24-JUN-1999; 99US-0140354.  
XX KW 28-JUN-1999; 99US-0140695.  
XX XX 29-JUN-1999; 99US-0140823.  
XX OS 29-JUN-1999; 99US-0140991.  
XX XX 30-JUN-1999; 99US-0141287.  
XX PN 01-JUL-1999; 99US-0141842.  
XX XX 01-JUL-1999; 99US-0142154.  
XX XX 02-JUL-1999; 99US-0142055.  
XX PD 06-JUL-1999; 99US-0142390.  
XX XX 08-JUL-1999; 99US-0142803.  
XX PF 09-JUL-1999; 99US-0142920.  
XX XX 12-JUL-1999; 99US-0142977.  
XX XX 13-JUL-1999; 99US-0143542.  
XX XX 14-JUL-1999; 99US-0143624.  
XX XX 15-JUL-1999; 99US-0144005.  
XX XX 16-JUL-1999; 99US-0144085.  
XX XX 16-JUL-1999; 99US-0144086.  
XX XX 19-JUL-1999; 99US-0144325.  
XX XX 19-JUL-1999; 99US-0144331.  
XX XX 19-JUL-1999; 99US-0144332.  
XX XX 19-JUL-1999; 99US-0144333.  
XX XX 19-JUL-1999; 99US-0144334.  
XX XX 19-JUL-1999; 99US-0144335.  
XX XX 20-JUL-1999; 99US-0144352.  
XX XX 20-JUL-1999; 99US-0144632.  
XX XX 20-JUL-1999; 99US-0144884.  
XX XX 21-JUL-1999; 99US-0144814.  
XX XX 21-JUL-1999; 99US-0145086.  
XX XX 21-JUL-1999; 99US-0145088.  
XX XX 22-JUL-1999; 99US-0145085.  
XX XX 22-JUL-1999; 99US-0145087.  
XX XX 22-JUL-1999; 99US-0145089.  
XX XX 22-JUL-1999; 99US-0145192.  
XX XX 23-JUL-1999; 99US-0145145.  
XX XX 23-JUL-1999; 99US-0145218.  
XX XX 23-JUL-1999; 99US-0145224.  
XX XX 26-JUL-1999; 99US-0145276.  
XX XX 27-JUL-1999; 99US-0145913.  
XX XX 27-JUL-1999; 99US-0145918.  
XX XX 27-JUL-1999; 99US-0145919.  
XX XX 28-JUL-1999; 99US-0145951.  
XX XX 02-AUG-1999; 99US-0146386.  
XX XX 02-AUG-1999; 99US-0146388.  
XX XX 02-AUG-1999; 99US-0146389.  
XX XX 03-AUG-1999; 99US-0147038.  
XX XX 04-AUG-1999; 99US-0147204.  
XX XX 04-AUG-1999; 99US-0147302.  
XX XX 05-AUG-1999; 99US-0147192.  
XX XX 05-AUG-1999; 99US-0147260.  
XX XX 06-AUG-1999; 99US-0147303.  
XX XX 06-AUG-1999; 99US-0147416.  
XX XX 09-AUG-1999; 99US-0147493.  
XX XX 09-AUG-1999; 99US-0147935.  
XX XX 10-AUG-1999; 99US-0148171.  
XX XX 11-AUG-1999; 99US-0148319.  
XX XX 12-AUG-1999; 99US-0148341.  
XX XX 13-AUG-1999; 99US-0148565.  
XX XX 16-AUG-1999; 99US-0148684.  
XX XX 17-AUG-1999; 99US-0149368.  
XX XX 17-AUG-1999; 99US-0149175.  
XX XX 18-AUG-1999; 99US-0149426.  
XX XX 20-AUG-1999; 99US-0149722.  
XX XX 20-AUG-1999; 99US-0149723.  
XX XX 20-AUG-1999; 99US-0149929.  
XX XX 23-AUG-1999; 99US-0149902.  
XX XX 23-AUG-1999; 99US-0149930.

18-OCT-2000 (first entry)  
Zea mays protein fragment SEQ ID NO: 43274.  
Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence; corn.  
Zea mays subsp. mays.  
EP1033405-A2.  
06-SEP-2000.  
25-FEB-2000; 2000EP-0301439.  
25-FEB-1999; 99US-0121825.  
05-MAR-1999; 99US-0123180.  
09-MAR-1999; 99US-0123548.  
23-MAR-1999; 99US-0125788.  
25-MAR-1999; 99US-0126264.  
29-MAR-1999; 99US-0126785.  
01-APR-1999; 99US-0127462.  
06-APR-1999; 99US-0128234.  
08-APR-1999; 99US-0128714.  
16-APR-1999; 99US-0129845.  
19-APR-1999; 99US-0130077.  
21-APR-1999; 99US-0130449.  
23-APR-1999; 99US-0130510.  
23-APR-1999; 99US-0130891.  
28-APR-1999; 99US-0131449.  
30-APR-1999; 99US-0132048.  
30-APR-1999; 99US-0132407.  
04-MAY-1999; 99US-0132484.  
05-MAY-1999; 99US-0132485.  
06-MAY-1999; 99US-0132486.  
06-MAY-1999; 99US-0132487.  
07-MAY-1999; 99US-0132863.  
11-MAY-1999; 99US-0134256.  
14-MAY-1999; 99US-0134218.  
14-MAY-1999; 99US-0134219.  
14-MAY-1999; 99US-0134221.  
14-MAY-1999; 99US-0134370.  
18-MAY-1999; 99US-0134768.  
19-MAY-1999; 99US-0134941.  
20-MAY-1999; 99US-0135124.  
21-MAY-1999; 99US-0135353.  
24-MAY-1999; 99US-0135629.  
25-MAY-1999; 99US-0136021.  
27-MAY-1999; 99US-0136392.  
28-MAY-1999; 99US-0136782.  
01-JUN-1999; 99US-0137222.  
03-JUN-1999; 99US-0137528.  
04-JUN-1999; 99US-0137502.  
07-JUN-1999; 99US-0137724.  
08-JUN-1999; 99US-0138094.  
10-JUN-1999; 99US-0138540.  
10-JUN-1999; 99US-0138847.  
14-JUN-1999; 99US-0139119.  
16-JUN-1999; 99US-0139452.  
16-JUN-1999; 99US-0139453.  
17-JUN-1999; 99US-0139492.  
18-JUN-1999; 99US-0139454.  
18-JUN-1999; 99US-0139455.  
18-JUN-1999; 99US-0139456.  
18-JUN-1999; 99US-0139457.  
18-JUN-1999; 99US-0139458.  
18-JUN-1999; 99US-0139459.  
18-JUN-1999; 99US-0139460.  
18-JUN-1999; 99US-0139461.

PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 12.1%; Score 84.5; DB 21; Length 361;  
Best Local Similarity 24.8%; Pred. No. 2.8;  
Matches 32; Conservative 16; Mismatches 46; Indels 35; Gaps 5;

QY 15 LPVYKGELEKGYQPDGWEISGFEGKK-----DAGVYNLS-----KDTFKPVPFKKI 61  
DB 193 IPVYK-----DNDSTFTVKRISDLQASKESPARAVASDKGTFVPSSSSQEKDPVAKTILNKV 248  
QY 62 EKKKEEENKPTFDVSKKKDNQVN-----HSQLNESHKKE-----DLOREH 103  
DB 249 KEKSDDEGKPLPAEFDGKENPVANSKNDKSDSGVDRSRERSRGRERDARGDRSDRDSR 308  
QY 104 SKSDSTKD 112  
DB 309 GRESDRERD 317

RESULT 39  
AAB35803

ID AAB35803 standard; Protein; 390 AA.  
XX AC AAB35803;  
XX DT 23-FEB-2001 (first entry)  
XX DE Protein involved in cell cycle regulation SEQ ID 38.  
XX KW Cell cycle regulation; corn; transgenic plant; cyclin; maize; soybean;  
KW cyclin-dependent kinase; sunflower; sorghum; canola; wheat; alfalfa;  
XX OS cotton; rice; barley; millet.  
XX Zea mays.  
XX PN WO200065040-A2.  
XX PD 02-NOV-2000.  
XX PF 13-APR-2000; 2000WO-US09975.  
XX PR 22-APR-1999; 99US-0130849.  
XX (PION-) PIONEER HI-BRED INT INC.  
XX Helentjaris TG, Habben JE, Sun Y;  
XX WPI; 2000-687333/67.  
XX DR N-PSDB; AAC83110.  
XX PT Nucleic acids useful for producing transgenic plants, preferably maize,  
PT with increased cell cycle gene activity, preferably activity of cyclin  
PT and/or cyclin-dependent kinase -  
XX Claim 16; Page 111-112; 122pp; English.

XX Polynucleotide sequences AAC831101 - AAC83113 encode proteins AAB35794 -  
AAC835806 which are involved in regulating the cell cycle. The protein and  
DNA sequences have been isolated from Zea mays (corn), and the invention  
also includes oligonucleotides AAC83114 - AAC83139 which are related to  
the cell cycle polynucleotides. The cell cycle polynucleotide sequences  
are useful for producing transgenic plants such as maize, soybean,  
CC sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley and  
CC millet with increased levels of cell cycle gene activity, such as  
CC activity of cyclin and cyclin-dependent kinases. The DNA sequences are  
CC also useful as probes for detecting deficiencies in the level of mRNA in  
CC screening for desired transgenic plants, for detecting mutations in the  
CC gene, for monitoring upregulation of expression or changes in enzyme  
CC activity in screening assays of compounds, for detecting any number of  
CC allelic variants, orthologs or paralogues of the gene, and site-directed  
CC mutagenesis in eukaryotic cells. The DNA sequences are also useful for  
CC recombinant expression of the encoded polypeptides and as immunogens for  
CC preparing and screening antibodies. A transgenic plant comprising an  
CC expression cassette including a cell cycle regulatory gene is useful for  
CC assaying enzyme agonists and antagonists, and as immunogens or antigens  
CC to obtain antibodies. The antibodies are useful in assaying expression  
CC levels of cell cycle regulatory proteins, for identifying and isolating  
CC nucleic acids from expression libraries, for identifying homologues of  
CC polypeptides from other species, and for purification of the proteins.

XX Sequence 390 AA;

Query Match 12.1%; Score 84.5; DB 21; Length 390;  
Best Local Similarity 24.8%; Pred. No. 3.1;  
Matches 32; Conservative 16; Mismatches 46; Indels 35; Gaps 5;

QY 15 LPVYKGELEKGYQPDGWEISGFEGKK-----DAGVYNLS-----KDTFKPVPFKKI 61  
DB 248 IPVYK-----DNDSTFTVKRISDLQASKESPARAVASDKGTFVPSSSSQEKDPVAKTILNKV 303  
QY 62 EKKKEEENKPTFDVSKKKDNQVN-----HSQLNESHKKE-----DLOREH 103  
DB 304 KEKSDDEGKPLPAEFDGKENPVANSKNDKSDSGVDRSRERSRGRERDARGDRSDRDSR 363

QY 104 SOKSDSTKD 112  
Db 364 GRESRD 372  
RESULT 40  
AAG35426  
ID AAG35426 standard; Protein; 416 AA.  
XX  
AC AAG35426;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Zea mays protein fragment SEQ ID NO: 43273.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence; corn.  
XX  
OS Zea mays subsp. mays.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 12.1%; Score 84.5; DB 21; Length 416;  
Best Local Similarity 24.8%; Pred. No. 3.4;  
Matches 32; Conservative 16; Mismatches 46; Indels 35; Gaps 5;  
QY 15 LPVYKELEKGVQFDWEISGPEGKK-----DAGYVINLS-----KDTFKPVFKKI 61  
DB 248 IPVYK----DNDSFTVKRISDLQASKESPARAVADKGTVPSSSQEKDPVAKTILNKV 303  
QY 62 EEKKEENKPTFDVSKKKNDPNQW-----HSOLNESHKKE-----DLQREEH 103  
DB 304 KEKSDDEGLPLAEFGCKENPVANSKNKDSGVDSDRERERGRGERDARGDSDRDR 363

QY 104 SQKSPSTKD 112  
DB 364 GRESDRERD 372

Search completed: February 10, 2004, 10:53:52  
Job time : 31.6275 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:45 ; Search time 9.67557 Seconds  
(without alignments)  
1331.870 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSSSTIVSEEDFILPVYK.....ATVLKKNISSKSTNNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	696	100.0	2140	2 F95074	serine proteinase,
2	693	99.6	2144	2 A37942	metalloproteinase,
3	110	15.8	558	2 T18467	hypothetical prote
4	108.5	15.6	1038	2 J05497	claustrin - chicke
5	99	14.2	211	2 T25911	hypothetical prote
6	98	14.1	665	2 B71609	hypothetical prote
7	97.5	14.0	1345	2 S46817	hypothetical prote
8	95	13.6	348	2 I37271	cyclicin II - human
9	95	13.6	622	2 A90570	lipoprotein [impor
10	94.5	13.6	312	2 G81339	probable membrane
11	94	13.5	210	2 T28771	hypothetical prote
12	94	13.5	535	2 T37189	hypothetical prote
13	93	13.4	1397	2 T10466	DNA topoisomerase
14	92.5	13.3	219	2 B72291	hypothetical prote
15	92.5	13.3	325	2 T18283	hypothetical prote
16	91	13.1	253	2 T32879	hypothetical prote
17	89.5	12.9	508	2 B81594	hypothetical prote
18	89.5	12.9	508	2 C72074	hypothetical prote
19	89.5	12.9	508	2 E86549	hypothetical prote
20	89	12.8	528	2 E96795	unknown protein F2
21	89	12.8	1888	2 T39009	hypothetical prote
22	89	12.8	3724	2 T18427	hypothetical prote
23	88.5	12.7	301	2 T33068	hypothetical prote
24	88.5	12.7	371	2 A71683	hypothetical prote
25	88.5	12.7	385	2 T20410	hypothetical prote
26	88.5	12.7	540	2 D86432	hypothetical prote
27	88.5	12.7	644	2 T47835	hypothetical prote
28	88.5	12.7	762	2 G88436	protein T04A8.13 [
29	88.5	12.7	791	2 T24435	hypothetical prote

ALIGNMENTS

RESULT 1

F95074

serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: F95074

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,

son, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: F95074

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2140 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK74791.1; PID:g14972117; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0641

Query Match 100.0%; Score 696; DB 2; Length 2140;

Best Local Similarity 100.0%; Pred. No. 2.2e-46;

Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGADAGYVNLKSDTFIKPVFKK 60

Db 1973 KEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKGADAGYVNLKSDTFIKPVFKK 2032

Qy 61 IEKKKEENKPTFDVSKKONPQVNHSQLNESHKRDLOREHSQKSDTKDVTATVLDK 120

Db 2033 IEKKKEENKPTFDVSKKONPQVNHSQLNESHKRDLOREHSQKSDTKDVTATVLDK 2092

Qy 121 NNISKSSTNNPNK 134

Db 2093 NNISKSSTNNPNK 2106

RESULT 2

A97942

metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C>Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001

C:Accession: A97942

R:Hoekings, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

Y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: A97942  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-2144 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99365.1; PID:g15458138; GSPDB:GN00174  
C:Genetics:  
A:Gene: pTtA  
C:Keywords: hydrolase; serine proteinase

Query Match 99.6%; Score 693; DB 2; Length 2144;  
Best Local Similarity 99.3%; Pred. No. 3.8e-46;  
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KEMSTTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFK 60  
|||||  
DB 1977 KEMSTTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFK 2036  
|||||

QY 61 IEKKEEENKPTFDVSKKNDPQVNHSQLNESHREDLQREHSQKSDSTKDVATVLDK 120  
|||||  
DB 2037 IEKKEEENKPTFDVSKKNDPQVNHSQLNESHREDLQREHSQKSDSTKDVATVLDK 2096  
|||||

QY 121 NNISKSTTNPNK 134  
|||||  
DB 2097 NNISKSTTNPNK 2110  
|||||

RESULT 3  
T18467  
hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2002  
C:Accession: T18467

R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, November 1998  
A:Reference number: Z18937

A:Accession: T18467  
A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-558 <LAW>

A:Cross-references: EMBL:AL008970; NID:e1407852; PIDN:CAA15610.1

C:Genetics:

A:Map position: 3

A:Introns: 84/1; 160/1

A:Note: C0465c

Query Match 15.8%; Score 110; DB 2; Length 558;  
Best Local Similarity 29.3%; Pred. No. 0.31;  
Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 7 IVSEEDFILPVY-----KGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPV 57  
|||  
DB 60 ILGFEDDILYEYCISQLKQSKK--KADGEEDKYLNAKKLKLNLGTFIGNKKSDFIBEL 117  
|||

QY 58 FKKI--EKKER-----ENKPTFDVSK-KKDNQVNHSQLNE-----SHRK 95  
|||  
DB 118 LELLNEEKKEEHIADTLNENK-TNDIRKVRNENINENVYNNKDSNKDKKEHVSHQN 176  
|||

QY 96 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISSKSTTN 130  
|||  
DB 177 EHNINNVNLKKEKEYTDIQRKKRHSLSQKSDSYKKRPFNKRKTSIER-SLSNRYDE 235  
|||

QY 131 NPNK 134  
|||  
DB 236 KTNK 239

RESULT 4  
JC5497  
clausstrin - chicken  
N:Alternate names: keratan sulfate proteoglycan  
C:Species: Gallus gallus (chicken)  
C:Date: 07-Jul-1997 #sequence\_revision 12-Sep-1997 #text\_change 21-Jul-2000  
C:Accession: JC5497; PC4334; S37561

R:Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A:Title: Clausstrin, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A:Reference number: JC5497; MUID:94157526; PMID:7906711

A:Accession: JC5497

A:Molecule type: mRNA

A:Residues: 1-1038 <BUR1>

A:Cross-references: EMBL:X67778; NID:g406318; PIDN:CAA47988.1; PID:g406319

A:Accession: PC4334

A:Molecule type: protein

A:Residues: 79-83;299-412;485-502 <BUR2>

A:Experimental source: brain

A:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nerve

C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate

F:267-270/Region: cell attachment (R-G-D) motif

F:112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 15.6%; Score 108.5; DB 2; Length 1038;

Best Local Similarity 28.6%; Pred. No. 0.83;

Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

QY 1 KEMSTTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINLSKDTFIKPVFK- 59  
|||  
DB 595 KPETKTIVAEDV-----TTKEEQLGKSETSEKQASEKQDVKPKVTKESVKKEVKA 646  
|||

QY 60 KIEKKEEENKPTFDVSKKNDPQVNHSQLNESHREDLQRE-----EHSQKSDSTKDV 113  
|||||  
DB 647 KPEEKDKDEKPKKEVSKKEEKLPI---KKEPKKEDIKKKEVKEVKEEKEKKEAKVE 702  
|||||

#### RESULT 5

T25911

hypothetical protein T23B3.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T25911

R:Maggi, L.; Le, T.

submitted to the EMBL Data Library, February 1997

A:Description: The sequence of C. elegans cosmid T23B3.

A:Reference number: Z20109

A:Accession: T25911

A:Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-211 <MAG>

A:Cross-references: EMBL:U88309; PIDN:AAB42334.1; GSPDB:GN00019; CESP:T23B3.5

A:Experimental source: strain Bristol N2; clone T23B3

C:Genetics:

A:Gene: CESP:T23B3.5

A:Map position: 1

A:Introns: 30/2; 200/3

Query Match 14.2%; Score 99; DB 2; Length 211;

Best Local Similarity 29.4%; Pred. No. 0.76;

Matches 30; Conservative 17; Mismatches 51; Indels 4; Gaps 1;

QY 37 EGKGDAGYVINLSKDTFIKPVFKIEKKEEENKPTFDVSKKNDPQVNH-----SQLNES 92  
|||  
DB 69 EGKGDGKKEKKGDKGKKEEKKDEKDDKKDEKDDKKDEKDDKKDEKDDKKDEKDD 128  
|||

QY 93 HRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 134  
|||

DB 129 EKKDDKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEK 170  
|||

#### RESULT 6

B71609

hypothetical protein PFB0680w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C:Accession: B71609

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;

Partea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.





A;Accession: G81339  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-312 <PAR>  
A;Cross-references: GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CAB72966.1; PID:G696814  
A;Experimental source: serotype O2, strain NCTC 11168  
C;Genetics:  
A;Gene: Cj0692c

Query Match 13.6%; Score 94.5; DB 2; Length 312;  
Best Local Similarity 25.7%; Pred. No. 2.7;  
Matches 38; Conservative 30; Mismatches 55; Indels 25; Gaps 7;

QY 8 VSEDFILPVYK-----GELEKGYQFDGWEISGPEGKDGAGVYINL---SKDTFIKP 56  
DB 38 ISSDDILRRFKKTPNKFLELDEYSEKTKSNYILKED---LINVKLEEKSLAKK 94  
QY 57 VFKEEKEEENKPT---FDVSKKKNP---QVNHSQLNESHKEDLQREEHKOKSDS 109  
DB 95 IFSKWKERKEENKTKNKNFLFSRKKAIEKNIQTKIQTKSNQATTTQKQEKELTNS 154  
QY 110 TKDV--TATVLDKNNISSK--STNNPN 133  
DB 155 IEKIQTETKIQKPLIEKLDVKNQPN 182

RESULT 11  
T28771  
hypothetical protein E03H12.5 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T28771  
R;Neelson, J.; Wohldmann, P.; Sansone, J.  
A;Description: The sequence of C. elegans cosmid E03H12.  
A;Reference number: Z20520  
A;Accession: T28771  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-210 <NEU>  
A;Cross-references: EMBL:AF000299; PIDN:AAC47980.1; GSPDB:GN00022; CBSP:E03H12.5  
A;Experimental source: strain Bristol N2; clone E03H12  
C;Genetics:  
A;Gene: CBSP:E03H12.5  
A;Map position: 4  
A;Introns: 30/2; 201/3

Query Match 13.5%; Score 94; DB 2; Length 210;  
Best Local Similarity 28.6%; Pred. No. 1.9;  
Matches 28; Conservative 18; Mismatches 52; Indels 0; Gaps 0;

QY 37 EGKDGAGVYINLSKDTFIKPVFKTEKEEENKPTFDVSKKKNPQVNHSQLNESHKRE 96  
DB 69 EGEKDGKDKKSEKKGKDEKKEKDEKKGKDEKDKDEKDKDEKDKDEKDEKDEKDE 128  
QY 97 DLQREEHKOKSDTKDVATVLDKNNISSKSTNNPNK 134  
DB 129 EKKDKDKDEKDEKDEKDEKSK 166

RESULT 12  
T37189  
hypothetical protein C02H7.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000  
C;Accession: T37189  
R;Leimac, D.; Minx, M.  
A;Description: The sequence of C. elegans cosmid C02H7.  
A;Reference number: Z20523  
A;Accession: T37189  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA

A;Accession: G81339  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-535 <LEI>  
A;Cross-references: EMBL:U49945; PIDN:AAC47924.1; GSPDB:GN00029; CBSP:C02H7.1  
A;Experimental source: strain Bristol N2; clone C02H7  
C;Genetics:  
A;Gene: CBSP:C02H7.1  
A;Map position: X  
A;Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 13.5%; Score 94; DB 2; Length 535;  
Best Local Similarity 23.4%; Pred. No. 5.3;  
Matches 33; Conservative 26; Mismatches 56; Indels 26; Gaps 4;

QY 4 SSTVSEDFILPVYKGELEKGYQFDGWEISGPEGKDGAGVYINLSKDTFIKPVFKTEE 63  
DB 86 AAKLISGKD-----AETNKMQLGNTATNSFRNGTG-----BEKKKKVKKEDK 132  
QY 64 KKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREEHKOKSDSTK-----DV 113  
DB 133 KGDEEEKST---TKRSSKKEEKEKSAEKEKKKKSSSSSKERHKSSDRSSEK 189  
QY 114 TATVLDKNNISSKSTNNPNK 134  
DB 190 SSEKSSKEKKKSTTDEKPK 210

RESULT 13  
T10466  
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - malaria parasite (Plasmodium falciparum)  
C;Species: Plasmodium falciparum  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000  
C;Accession: T10466  
R;Cheesman, S.J.  
A;Description: submitted to the EMBL Data Library, September 1995  
A;Reference number: Z17031  
A;Accession: T10466  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1397 <CHE>  
A;Cross-references: EMBL:X79345; NID:G994807  
C;Genetics:  
A;Gene: TopoII  
A;Map position: 14  
C;Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hyd  
C;Keywords: ATP; DNA binding; isomerase; nucleus

Query Match 13.4%; Score 93; DB 2; Length 1397;  
Best Local Similarity 24.6%; Pred. No. 19;  
Matches 35; Conservative 33; Mismatches 48; Indels 26; Gaps 6;

QY 8 VSEDFIL--PVYKGELEKGYQFDGWEISGPEGKDGAGVYINLSKDTFIKPVFKTEE-- 63  
DB 1141 VKDYDYLLSMFIFSLTLEK---VEDLLTQLKEKEKELEILNITVTETNWLKDIKVEEAI 1197  
QY 64 -----KKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSHQ---KSDSTK 111  
DB 1198 EFQNVLSNREESNK--FKVARQKQ-----GPSMKKKKKKKLSSDESGDTSDES 1250  
QY 112 DVTATVLDKNNISSKSTNNPN 133  
DB 1251 FLVNTLNKKNTKTTSSNN 1272

RESULT 14  
B72291  
hypothetical protein - Thermotoga maritima (strain MSB8)  
C;Species: Thermotoga maritima  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: B72291  
R;Neelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
A;Description: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
Nature 399, 323-329, 1999  
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq

A;Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: B72291

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-219 <ARN>

A;Cross-references: GB:AE001771; GB:AE000512; NID:94981678; PIDN:AAD36218.1; PID:9498168

A;Experimental source: strain MSB8

C;Genetics:

A;Map position: 1

A;Introns: 41/1

Query Match 13.3%; Score 92.5; DB 2; Length 219;  
Best Local Similarity 28.1%; Pred. No. 2.5;  
Matches 36; Conservative 24; Mismatches 35; Indels 33; Gaps 8;

Qy 20 GELEKGYQ--FDGWEISG-----FEKGDAGYVIN-LSKDTFTKPVFKKIEEKEENK 70

Db 89 GEESVSRDLFAGVGVRGTPTFFFKGEGGLGYLPGVDKDNFIK-ILKYVAQLKED-- 145

Qy 71 PTFDVSKKONPVNHSQLNESHKED----LQREHSQKSDS-----TKD 112

Db 146 --FOTYLKQDPFVGEPLIIEIF-KEDADFVLEKDNANVKDVTVPNEVRDRIYVTDSPD 202

Qy 113 VTATVLOK 120

Db 203 VAKTLOEK 210

RESULT 15

T18283

hypothetical protein G5 - slime mold (Dictyostelium discoideum)

C;Species: Dictyostelium discoideum

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C;Accession: T18283

R;Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh

Genetics 148, 1117-1125, 1998

A;Title: Dictyostelium discoideum nuclear plasmid, ddp5 is a chimera related to the Ddp1

A;Reference number: 214684; MUID:98198836; PMID:9539429

A;Accession: T18283

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-325 <RIE>

A;Cross-references: EMBL:U00796; NID:92702254; PID:g2702258; PIDN:AAAC18634.1

C;Genetics:

A;Introns: 85/1

Query Match 13.3%; Score 92.5; DB 2; Length 325;  
Best Local Similarity 23.2%; Pred. No. 4;  
Matches 38; Conservative 23; Mismatches 42; Indels 61; Gaps 8;

Qy 8 VSEEDFILPVYK-GEL--EKGYQFDGWEISGFEGK-----KDAGYV-----I 46

Db 72 IGDEKLFRIKNGELIVLNELEFDNFHFK--EGKHLKSKMFMHVKDSGYATNEBIEI 129

Qy 47 NLSKDTFTKPV-----FKK-----IIEKKEENKPTFDV 75

Db 130 FLESTLCKEITATQTKNSYKKNLKNLPEEEEEEEEEEEEEEEEEEEQEEVEKPTISE 189

Qy 76 SKKNDONPVNHSQLNESHKKE-----DLQREHSQKSDSTKD 112

Db 190 EEBEETPAVSEEEKEEEEEETPAVSEEEKEEEEEQEEDEKED 233

RESULT 16

T32879

hypothetical protein C17F3.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C;Accession: T32879

R;Gattung, S.; Scheet, P.

submitted to the EMBL Data Library, January 1998

A;Description: The sequence of C. elegans cosmid C17F3.

A;Reference number: Z21240

A;Accession: T32879

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-253 <GAT>

A;Cross-references: EMBL:AF043692; PIDN:AAB97531.1; GSPDB:GN00019; CESP:C17F3.3

A;Experimental source: strain Bristol N2; clone C17F3

C;Genetics:

A;Gene: CESP:C17F3.3

A;Map position: 1

A;Introns: 41/1

Query Match 13.1%; Score 91; DB 2; Length 253;  
Best Local Similarity 32.6%; Pred. No. 3.9;  
Matches 31; Conservative 7; Mismatches 33; Indels 24; Gaps 3;

Qy 56 PVFKKTEKEKE---EENKPTFDVSKK-----KDNQPVNHSQLNESHK----- 94

Db 61 PVAPKVEEKEEKEEKKADDEKKEEKKTEEDKDKSKKTEEDKDKSVKKTQETKSERKDKK 120

Qy 95 ---KEDLQREHSQKSDSTKDVVTATVLDKNNISSK 126

Db 121 DEKDEKKEEKEKSKDEEKKDKVKKDEKK 155

RESULT 17

B81594

hypothetical protein CP0281 [imported] - Chlamydomophila pneumoniae (strain AR39)

C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 11-May-2000

C;Accession: B81594

R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,

C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A;Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A;Reference number: A81500; MUID:20150255; PMID:10684935

A;Accession: B81594

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-508 <REA>

A;Cross-references: GB:AE002189; GB:AE002161; NID:g7189205; PIDN:AAF38139.1; PID:g718920

A;Experimental source: strain AR39, HL cells

C;Genetics:

A;Gene: CP0281

Query Match 12.9%; Score 89.5; DB 2; Length 508;  
Best Local Similarity 24.5%; Pred. No. 11;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

Qy 54 IKPVFKKIEEKEEENKPTFD-----VSKKNDONPVNHSQLNESHK 95

Db 92 VKGVFKKTPQARPEVSPRLPSPHVQHQQLPGLGFRDRIQKRSENPEADLGKMKRSYSD 151

Qy 96 EDLQREHSQKSDSTKDVVTATVLDKNNISSKTT 129

Db 152 GDLDRVGHSDNEDSTEDSR-----EGGEPSPSKSS 182

RESULT 18

C72074

hypothetical protein - Chlamydomophila pneumoniae (strain CWL029)

C;Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000

C;Accession: C72074

R;Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;

Nature Genet. 21, 385-389, 1999

A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A;Reference number: A72000; MUID:99206606; PMID:10192388

A;Accession: C72074

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-508 <ARN>

A;Cross-references: GB:AE001632; GB:AE001363; NID:g4376755; PIDN:AAD18613.1; PID:g437675

A;Experimental source: strain CWL029

C;Genetics:

A;Gene: CPN0473

Query Match      12.9%; Score 89.5; DB 2; Length 508;  
Best Local Similarity    24.5%; Pred. No. 11;  
Matches         23; Conservative    17; Mismatches    33; Indels    21; Gaps                  2;

QY    54 IKPVFKIEEKEENKPTFD-----VSKKKDNPNVNHSQNLSHRK 95  
| | | | : : : :  
Db    92 VGVFVKTPQARPEVSPPRLSHVGORQLPGLGFRDRIQRSENPEADLGNKRYSYSD 151  
| | | | : : : :  
QY    96 EDLOREHQSXSSTKDVTATVLDDKNISSSKT 129  
| | | | : : : :  
Db    152 GLDLRVGHDSNEDETSRS---EGGEPSKSXS 182  
| | | | : : : :

RESULT 19  
E86549  
hypothetical protein CPJ0473 [imported] - Chlamydophila pneumoniae (strain JI38)  
C;Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
R;Accession: E86549  
R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise,  
Nucleic Acids Res. 28, 2311-2314, 2000  
A;Title: Comparison of whole genome sequences of chlamydia pneumoniae JI38.  
A;Reference number: A86491; PMID:20330349; PMID:10871362  
A;Accession: E86549  
A>Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-508 <STO>  
A;Cross-references: GB:A000008; NID:g9978843; PIDN:BAA98679.1; MSPDB:GN00142  
A;Experimental source: strain JI38  
C;Genetics:  
A;Gene: CPJ0473

Query Match      12.9%; Score 89.5; DB 2; Length 508;  
Best Local Similarity    24.5%; Pred. No. 11;  
Matches         23; Conservative    17; Mismatches    33; Indels    21; Gaps                  2;

QY    54 IKPVFKIEEKEENKPTFD-----VSKKKDNPNVNHSQNLSHRK 95  
| | | | : : : :  
Db    92 VGVFVKTPQARPEVSPPRLSHVGORQLPGLGFRDRIQRSENPEADLGNKRYSYSD 151  
| | | | : : ~~~~~~  
QY    96 EDLOREHQSXSSTKDVTATVLDDKNISSSKT 129  
| | | | : : | | : :  
Db    152 GLDLRVGHDSNEDETSRS---EGGEPSKSXS 182  
| | | | : : | | : :

RESULT 20  
E96795  
unknown protein F28016.8 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
R;Accession: E96795  
R;Theologis, A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;  
ansen, N.P.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.-  
C.A.; Li, J.H.; Liu, Y.; Lin, X.; Liu, S.X.; Lu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shimn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; PMID:21016719; PMID:11130712  
A;Accession: E96795  
A>Status: preliminary  
A:Molecule type: DNA  
A;Residues: 1-528 <STO>  
A;Cross-references: GB:AE005173; NID:G61433888; PIDN:AAF04434.1; MSPDB:GN00141  
C;Genetics:  
A;Gene: F28016.8  
A;Map position: 1

Query Match      12.8%; Score 89; DB 2; Length 528;  
Best Local Similarity    24.3%; Pred. No. 13;  
Matches         35; Conservative    22; Mismatches    51; Indels    36; Gaps                  6;

QY    21 ELEKGYPDGW-----EISFECKDKAG----YYINLNKDTFIKPVEPKIEE 63  
| | | | : : | | : | | | : : : :  
Db    200 ELAYDFNFETGGAKVCCLCAACAVSFLGAQSGFGASPYVILLIEDTY--VWDGD 256  
| | | | : : | | : | | | : : | | : :  
QY    64 KEEBENKPTFDVVSKKK--DNPNVNHSQNLSHRKEDIQREEH-----SQXSDTKD- 112  
| | | | : : | | : | | | : : | | : :  
Db    257 RYSVDKIPIVDADELTSBPNGESNTNEEKIKISTEHLESLALNIQQOSDSTPTT 316  
| | | | : : | | : | | | : : | | : :  
QY    113 -----VTATVLDDKNISSKTTNN 131  
| | | | : : | | : | | | : : | | : :  
Db    317 MEEDVTVTEKTSSEDMKLILSON 340  
| | | | : : | | : | | | : : | | : :

RESULT 21  
T39009  
hypothetical protein SPAC6B12.02c - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomycetes pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000  
R;Accession: T39009  
R;Gentiles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.,  
submitted to the EMBL Data Library, August 1995  
A;Reference number: Z21815  
A;Accession: T39009  
A>Status: preliminary; translated from GB/EMBL/DDBU  
A:Molecule type: DNA  
A;Residues: 1-1888 <GEN>  
A;Cross-references: EMBL:Z98531; PIDN:CAB11064.1; MSPDB:GN00066; SPDB:SPAC6B12.02c  
A;Experimental source: strain 972h-; cosmid c6B12  
C;Genetics:  
A;Gene: SPDB:SPAC6B12.02c  
A;Map position: 1  
C;Superfamily: Schizosaccharomycetes hypothetical protein SPAC6B12.02c

Query Match      12.8%; Score 89; DB 2; Length 1888;  
Best Local Similarity    23.9%; Pred. No. 54;  
Matches         37; Conservative    29; Mismatches    59; Indels    30; Gaps                  7;

QY    4 SSTIVSEDFF-----ILPVY---KGELEGKQGPDGWELSGFE-----GKKDAG 43  
| | | | : | | | : | | : | | | : | | : | | :  
Db    390 SSLLTSENPFOLNVANASTIPVTRTTKMKN-RFYKYEVELPLDILESYGKKAPK 448  
| | | | : | | | : | | : | | | : | | : | | :  
QY    44 YYINLNKDTFTFKVPFKIEEKENKPTFDVSKKKDPNVNHSQNLSHRKEDLOREH 103  
| | | | : | | | : | | : | | | : | | : | | :  
Db    449 FLRFVARSSHIPP-KMIRRRQMDSKKYFSFKESDRQVIDOVLSDWYGKHVLQQSH 506  
| | | | : | | | : | | : | | | : | | : | | :

QY    104 SQKSDS-TKVDTATLDKN-----NISSSKTTNN 131  
| | | | : | | | : | | : | | | : | | : | | :  
Db    507 SYKXPDSKSVCGNIFSNVNSKKHVNLNAKAANN 541  
| | | | : | | | : | | : | | | : | | : | | :

RESULT 22  
TI8427  
hypothetical protein CO335c - malaria parasite (Plasmodium falci-parum)  
C;Species: Plasmodium falci-parum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
R;Accession: TI8427  
R;Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, August 1997  
A;Reference number: Z18935  
A;Accession: TI8427  
A>Status: preliminary; translated from GB/EMBL/DDBU  
A:Molecule type: DNA  
A;Residues: 1-3724 <LAU>  
A;Cross-references: EMBL:Z98547; NID:e1325376; PIDN:e1325379; PIDN:CAB11104.1  
C;Genetics:  
A;Introns: 307/1; 1545/2  
A>Note: CO335c

Query Match      12.8%; Score 89; DB 2; Length 3724;

Best Local Similarity 22.3%; Pred. No. 1.2e+02;  
Matches 29; Conservative 27; Mismatches 42; Indels 32; Gaps 5;  
Qy 6 TIVSEEDFILPVYKGELEKGYQFGWEISGFEKK-DAGYVINLSKDTFIKPVFKIEKK 65  
Db 1060 TNICDENNIQOINSEKKGVRISGTDN---ENKND-----MENKN 1097  
Qy 66 EEENKPTFDVSKKKNPQVNH-SQLNESHKEDLQREHSQKSDTKDVTATVLDKNIS 124  
Db 1098 DMEKKN--DMEKNDIEKNDMEKNDMEKNDMEKNDMEKNDMEKNDMEKNDME 1148  
Qy 125 SKSTNNPNK 134  
Db 1149 NKSDIENENK 1158

## RESULT 23

T33068

hypothetical protein C35E7.9 - *Caenorhabditis elegans*C/Species: *Caenorhabditis elegans*

C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C/Accession: T33068

R/Graves, T.; McDonald, R.

submitted to the EMBL Data Library, May 1998

A/Description: The sequence of *C. elegans* cosmid C35E7.

A/Reference number: Z21278

A/Accession: T33068

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-301 &lt;GRA&gt;

A/Cross-references: EMBL:AF067216; PIDN:AAC17524.1; GSPDB:GN00019; CESP:C35E7.9

A/Experimental source: strain Bristol N2; clone C35E7

C/Genetics:

A/Gene: CESP:C35E7.9

A/Map position: 1

A/Introns: 30/3; 193/1; 236/2

Query Match 12.7%; Score 88.5; DB 2; Length 301;  
Best Local Similarity 24.8%; Pred. No. 7.5;  
Matches 27; Conservative 17; Mismatches 38; Indels 27; Gaps 3;  
Qy 25 GYQFDGWEISGFEKK-----DAGYVINLSKDTFIKPVFKI-----EE 63  
Db 10 GITTAGWLAGCGGKKKDGKSGTASAAAPKADSKMKPPVENVKSKKSEKKEPKGKEEP 69  
Qy 64 KKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKD 112  
Db 70 KKEEKEKSKKSEKDDKK-----EAKKEDDKDEKDEKDDKD 112

## RESULT 24

A71683

hypothetical protein RP278 - *Rickettsia prowazekii*C/Species: *Rickettsia prowazekii*

C/Date: 21-Nov-1998 #sequence\_revision 21-Nov-1998 #text\_change 03-Nov-2000

C/Accession: A71683

R/Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichert-Ponten, T.; Alsmark, U.

Nature 396, 133-140, 1998

A/Title: The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.

A/Reference number: A71630; MUID:99039499; PMID:9823893

A/Accession: A71683

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-371 &lt;AND&gt;

A/Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14739.1; PID:g386083

A/Experimental source: strain Madrid E

C/Genetics:

A/Gene: RP278

Query Match 12.7%; Score 88.5; DB 2; Length 371;  
Best Local Similarity 23.1%; Pred. No. 9.5;  
Matches 34; Conservative 33; Mismatches 37; Indels 43; Gaps 9;

Qy 1 KEMSTIVSEEDFILPVYKGELEKGYQFGWEISGFEKK-DAGYVINLSKDTFIKPVFK 59  
Db 192 EQLNLTLINEE-----FRKNLEO-----EKKEGKKTQASEALNKK-----LKPIYK 234  
Qy 60 KIEKKEE-----ENKPTFDVSKKKNPQVNH-SQLNESHKEDLQREHSQKSDTK 111  
Db 235 QMDEEREELFKLAELLPQY-----AQANIDGHAKLYAKQYOTKIENDPNYKELEKQ 286  
Qy 112 DVTATVLDKNISKSTT-----NNPN 133  
Db 287 E-----IVSKIEYKSKSKNDIINPN 309

## RESULT 25

T20410

hypothetical protein E02A10.2 - *Caenorhabditis elegans*C/Species: *Caenorhabditis elegans*

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T20410

R/Thomas, K.

submitted to the EMBL Data Library, October 1996

A/Reference number: Z19271

A/Accession: T20410

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-385 &lt;WIL&gt;

A/Cross-references: EMBL:Z81053; PIDN:CAB02877.1; GSPDB:GN00023; CESP:E02A10.2

A/Experimental source: clone E02A10

C/Genetics:

A/Gene: CESP:E02A10.2

A/Map position: 5

A/Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

Query Match 12.7%; Score 88.5; DB 2; Length 385;  
Best Local Similarity 35.5%; Pred. No. 9.9;  
Matches 22; Conservative 14; Mismatches 19; Indels 7; Gaps 2;

Qy 59 KKEEKEEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVL 118  
Db 317 KKEEKEEKEEKE--EVEKKEE-----EKKDEPKKEEKEEKEEKEEVEEKEKVE 369

Qy 119 DK 120

Db 370 EK 371

## RESULT 26

D86432

hypothetical protein T518.14 - *Arabidopsis thaliana*C/Species: *Arabidopsis thaliana* (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C/Accession: D86432

R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Matzali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: D86432

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-540 &lt;STO&gt;

A/Cross-references: GB:AE005172; NID:g4587525; PIDN:AAD25756.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

Query Match 12.7%; Score 88.5; DB 2; Length 540;  
Best Local Similarity 24.4%; Pred. No. 14;

Matches 31; Conservative 25; Mismatches 54; Indels 17; Gaps 4;  
QY 2 ENSSTVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKK- 60  
DB 49 EKSAFKEESDFADLKESEKK-----ALSDLKSLKEAIVDN---TLLTKKKES 96  
QY 61 --IEKKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVL 118  
DB 97 SPMSKKKEEVKPEAEVEKKKE--EAAEKVEEKKSEAVVTEAPKAETVEAVVTESII 154  
QY 119 DKNNTSS 125  
DB 155 PKEEVT 161

RESULT 27  
T47835  
hypothetical protein T209.90 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C:Accession: T47835  
R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
submitted to the Protein Sequence Database, February 2000  
A:Reference number: Z24475  
A:Accession: T47835  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <NYA>  
A:Cross-references: EMBL:AL138658  
A:Experimental source: cultivar Columbia; BAC clone T209  
C:Genetics:  
A:Map position: 3  
A:Introns: 158/2; 329/3  
A:Note: T209.90

Query Match 12.7%; Score 88.5; DB 2; Length 644;  
Best Local Similarity 24.6%; Pred. No. 18;  
Matches 32; Conservative 18; Mismatches 53; Indels 27; Gaps 3;  
QY 9 SEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKKIEKKEE 68  
DB 534 SVADFLKRIKNSPQKG-----ETTSKNQKNDGNV-----KKEND 570  
QY 69 NKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDS-----TKDVTATVLDKNNIS 124  
DB 571 HOKSDGNVKKNSKVPKRLRSSTGKKKVEVNNNSKSKRKQTKETAEVATGKRGRE 630  
QY 125 SKSTNNPNK 134  
DB 631 SGKDKQPRK 640

RESULT 28  
G88436  
protein T04A8.13 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
C:Accession: G88436  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/c\_elegans/ and www.sanger.ac.uk/Projects/C\_ele  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: G88436  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-762 <STO>  
A:Cross-references: GB:chr\_III; PIDN:CAA84732.1; PID:g3879347; GSPDB:GN000021; CESP:T04A8  
C:Genetics:  
A:Gene: T04A8.13  
A:Map position: 3

Query Match 12.7%; Score 88.5; DB 2; Length 762;  
Best Local Similarity 25.2%; Pred. No. 21;  
Matches 32; Conservative 25; Mismatches 65; Indels 5; Gaps 3;  
QY 9 SEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKKIEKKEE 67  
DB 14 SGKQIMKPGY--DKKEGLGMDQKEIVGDDKKDKKARKRKRLQDEFAE--LKDEEKDKE 69  
QY 68 ENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKS 127  
DB 70 EAEKEKNEKEKKEKEDGHEKKEDKEDKENENDEKKEKSKDDKKEESKEDKKEKTK 129  
QY 128 TTNNPNK 134  
DB 130 TEDNEKG 136

RESULT 29  
T24435  
hypothetical protein T04A8.13 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T24435  
R:Palmer, S.  
submitted to the EMBL Data Library, August 1994  
A:Reference number: Z19889  
A:Accession: T24435  
A>Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-791 <WIL>  
A:Cross-references: EMBL:Z35663; PIDN:CAA84732.2; GSPDB:GN000021; CESP:T04A8.13  
A:Experimental source: clone T04A8  
C:Genetics:  
A:Gene: CESP:T04A8.13  
A:Map position: 3  
A:Introns: 31/3; 212/1; 229/3; 331/3; 406/1; 472/3; 572/1; 651/2

Query Match 12.7%; Score 88.5; DB 2; Length 791;  
Best Local Similarity 25.2%; Pred. No. 22;  
Matches 32; Conservative 25; Mismatches 65; Indels 5; Gaps 3;  
QY 9 SEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKKIEKKEE 67  
DB 14 SGKQIMKPGY--DKKEGLGMDQKEIVGDDKKDKKARKRKRLQDEFAE--LKDEEKDKE 69  
QY 68 ENKPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISKS 127  
DB 70 EAEKEKNEKEKKEKEDGHEKKEDKEDKENENDEKKEKSKDDKKEESKEDKKEKTK 129  
QY 128 TTNNPNK 134  
DB 130 TEDNEKG 136

RESULT 30  
T28676  
rhoptry protein - Plasmodium yoelii (fragment)  
C:Species: Plasmodium yoelii  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 01-Dec-2000  
C:Accession: T28676; A45521  
R:Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.  
Mol. Biochem. Parasitol. 76, 329-332, 1996  
A:Title: Comparison of two members of a multigene family coding for high-molecular mass  
A:Reference number: Z20507; MUID:97077455; PMID:8920022  
A:Accession: T28676  
A>Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-2401 <SIN>  
A:Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AAB41263.1  
R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.  
Mol. Biochem. Parasitol. 42, 241-246, 1990  
A:Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co  
A:Reference number: A45521; MUID:91101660; PMID:2270106

```
A;CROSS-references: SGD:S0002232  
A;Map position: 4L  
C;Keywords: transmembrane protein  
F;69-85/Domain: transmembrane #status predicted <TMM>  
  
Query Match          12.6%; Score 88; DB 2; Length 700;  
Best Local Similarity 23.7%; Pred. No. 21;  
Matches      31; Conservative    24; Mismatches   40; Indels     36; Gaps       5;
```

QY            21 ELEGVGFQDGEWISGPEGKK-----DAGYVINL-SKDTFF---IK 55

       :|||:::||  
Db      447 DLEKGFR----ELSDLTHKYSEIINHESVISKLTVETKRADQKYFAAMESKDSILLIEIK 502

QY            56 PVFKKEEKEENKTFTFDVSCKKDHPVNHSQLNESHXRDEDLOREHSSQSSTKDVT A 115

       :||:||||:~::~:~::~:  
Db      503 TLSKSUSKNEL-----ILQLKSDRLQLQGHNHLKHLDLSQNERRLLIDSSKTETT L 555

QY            116 TVLDDKNINISK 126

       ::|||!:

Db          556 KIIDLNNITSK 566

       ::|||!

RESULT 33

A:IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus  
 C:Species: Haemophilus influenzae  
 A:Variety: strain HK715  
 C:Date: 04-Mar-1993 #sequence\_revision 19-Nov-1994 #text\_change 08-Dec-2000  
 C:Accession: A41859  
 R:Poulsen, K.; Reinholdt, J.; Kilian, M.  
 J. Bacteriol. 174, 2913-2921, 1992  
 A:Title: A comparative genetic study of serologically distinct Haemophilus infl  
 A:Reference number: A41859; PMID:92234949; PMID:1373717  
 A:Accession: A41859  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-1702 <FO>  
 A:Cross-references: GB:M87489; NID:G148906; PIDN:AAA24966.1; PID:G148907  
 A:Experimental source: strain HK715  
 A:Note: sequence extracted from NCBI backbone (NCBIP:97282)  
 C:Superfamily: IGA-specific metalloendopeptidase  
 C:Keywords: hydrolase; metalloproteinase

```
Query Match      12.6%; Score 87.5; DB 2; Length 1702;
Best Local Similarity 27.2%; Pred.No. 63;
Matches          25; Conservative 14; Mismatches 50; Indels 3; Gaps 1;

Qy    46 INLSKDTFIKPVPKIEEKEENKPTFDYVKKDNPVVHSQLNSHRKEDLOREHS - 104
      ||| : : : : : : : : : : : : : : : : : |
Db    1296 INTGSAITATETAEKGDKPOTEAATEDASOHKANTVDNSVANNESSPEPKRRRSI 1355
      ||| : : : : : : : : : : : : : : : : : |

Qy    105 --QKS DSTKVATATVLDKNNISSKTNTNPNK 134
      ||| : : : : : : : : : : : : : : : : : |
Db    1356 SQPQETSAAETTAASTDETTIANDSKRKSPNR 1387
      ||| : : : : : : : : : : : : : : : : : |
```

RESULT 34

T18477  
 hypothetical protein C0485w - malaria parasite (Plasmodium falciparum)  
 C/Species: Plasmodium falciparum  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2002  
 C/Accession: T18477  
 R/Lawson, D.; Bowman, S.; Barrell, B.  
 submitted to the EMBL Data Library, November 1998  
 A/Reference number: Z18937  
 A/Accession: T18477  
 A/Status: preliminary; translated from GE/EMBL/DBDJ  
 A/Molecule type: DNA  
 A/Residues: 1-2523 <LAW>  
 A/Cross-references: EMBL:AL008970; NID:el407852; PIDN:CAAL5620.1  
 C/Genetics:  
 A/Map position: 3  
 A/Introns: 148/3

A>Note: C0485W

Query Match 12.6%; Score 87.5; DB 2; Length 2523;  
Best Local Similarity 32.5%; Pred. No. 98;  
Matches 25; Conservative 22; Mismatches 21; Indels 9; Gaps 5;  
QY 59 KXIEKKEENKPTFDVSKKD-NPQVNHSQLNESHKEDL-OREEHQSQSDSTK-DVT 114  
Db 1795 KKKIKKDDINKKE-DINKKODINKKNDINKKDDINKKDDINKKDDINKKDDIN 1853  
QY 115 ATVLDKNNISSKSTNN 131  
Db 1854 ----KKNYNNNNNN 1866  
RESULT 35  
F71621  
Hypothetical protein PF0170w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C>Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: F71621  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Pertes, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: F71621  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-635 <GAR>  
A:Cross-references: GB:AE001377; GB:AE001362; NID:g3845111; PIDN:AAC71823.1; PID:g384511  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PF0170w

Query Match 12.5%; Score 87; DB 2; Length 635;  
Best Local Similarity 23.8%; Pred. No. 23;  
Matches 30; Conservative 20; Mismatches 52; Indels 24; Gaps 4;  
QY 22 LEKQYQFGWEISGFEKGDAGYVNLKSDTKFVFKVFKIEKKEENKPTFDVSKKDN 81  
Db 202 LQKYNQIDDEEDNETIRSDKLDIYSDSQSDIMWSSPNKEES-----MSSDNHN 256  
QY 82 POWNHSQLNESHKEDLQREHQSQSDTKVATVLD-----KNNISKS 127  
Db 257 KDINS-----DNQKNDINSSDHNM-NDSTNESTTTSLSTSNNTNRNKNRKNINNN 311  
QY 128 TTNPN 133  
Db 312 NNNNSN 317

RESULT 36  
QMSPI  
Microtubule-associated protein MAP1B - mouse  
N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei  
C:Species: Mus musculus (house mouse)  
C>Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 01-Sep-2000  
C:Accession: S07549; S44387; A33645  
R:Noble, M.; Lewis, S.A.; Cowan, N.J.  
J. Cell Biol. 109, 3367-3376, 1989  
A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contains  
A:Reference number: A33645; MUID:90094539; PMID:2480963  
A:Accession: S07549  
A:Molecule type: mRNA  
A:Residues: 1-2464 <NOB>  
A:Cross-references: EMBL:X51396; NID:952999; PIDN:CAA35761.1; PID:g53000  
R:Sanchez, C.; Padilla, R.; Paciucci, R.; Zabaia, J.C.; Avila, J.  
Arch. Biochem. Biophys. 310, 428-432, 1994  
A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.  
A:Reference number: S44387; MUID:94234720; PMID:8179328  
A:Accession: S44387  
A:Status: preliminary

A:Molecule type: protein  
A:Residues: 653-663, 'IC' <SAN>  
C:Superfamily: microtubule-associated protein MAP1B  
C:Keywords: microtubule binding; phosphoprotein; tandem repeat  
F:589-786/Domain: microtubule binding #status experimental <MTB>  
F:589-592, 639-642, 649-652, 655-658, 660-661, 674-677, 679-682, 683-686, 687-690, 691-696  
R-K-E/D-X  
F:1861-2064/Region: 17-residue repeats  
F:91-116, 351, 888, 1124, 1153, 1168, 1208, 1662, 1877, 1918, 2003, 2030, 2054, 2083/Binding site: ph  
F:147, 969, 1336, 1562, 1563, 1702, 1708, 1990, 2057, 2063, 2419/Binding site: phosphate (Thr) (co  
F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 12.5%; Score 87; DB 1; Length 2464;  
Best Local Similarity 28.4%; Pred. No. 1e+02;  
Matches 29; Conservative 16; Mismatches 37; Indels 20; Gaps 4;  
QY 48 LSKDTFKVPFK-KIEKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQRE----- 101  
Db 632 VTQKWKVKIEIKTKLEKKKEE-KPKKEVVKDKTLP---KDEKPKKEVKKKEIKKEI 686  
QY 102 -----EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 134  
Db 687 KKEERKELKKEVKKETPLKDAKKEVKKEEKEVKEKEPEKK 728

RESULT 37  
T28391  
ORF MSV230 hypothetical protein - Melanoplus sanguinipes entomopoxvirus  
C:Species: Melanoplus sanguinipes entomopoxvirus  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T28391  
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: Z20484; MUID:99102612; PMID:9847359  
A:Accession: T28391  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-670 <AFO>  
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97713.1; PID:g4049753  
C:Genetics:  
A>Note: MSV230

Query Match 12.4%; Score 86.5; DB 2; Length 670;  
Best Local Similarity 22.7%; Pred. No. 26;  
Matches 35; Conservative 33; Mismatches 65; Indels 21; Gaps 5;  
QY 1 KEMSTI-----VSEEDFILPVYKGELEKGY-----QFDGWEISGFEKGD 42  
Db 39 KQIISTLLKFNFDKTCMGVSEKVFQVLNNKSAKESKYSVDVSSIDESQSDSDSDS 98  
QY 43 GYVNLKSDTKFVKPKIE-EKKEENKPTFDVSKKKNPQVNHSQL-NESHKEDLQ 100  
Db 99 GWNIDESQSDSKVNINKLENESQSDSKVNIDESQSDS-KVNINKLENESQSDSKVN 157  
QY 101 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 134  
Db 158 IDESQSDSKVNIDESQSDSKVNIDESQSDSK 191

RESULT 38  
T49989  
Hypothetical protein F12B17.150 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49989  
R:Byvan, M.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, April 2000  
A:Reference number: Z25026  
A:Accession: T49989  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-792 <BEV>



```

RESULT 40
19006
nkyrin related protein C06C3.1 - Caenorhabditis elegans
;Contains: myosin-light-chain-phosphatase (EC 3.1.3.53)
;Species: Caenorhabditis elegans
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Mar-2000
;Accession: T19006; T22086
;Berks, M.
submitted to the EMBL Data Library, August 1994
;Reference number: Z19058

```

***This Page Blank (uspto)***

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	97.5	14.0	1345	1	YHQO YEAST	P38800 saccharomyc
2	95	13.6	348	1	CYL2 HUMAN	Q14093 homo sapien
3	93	13.4	1398	1	TOP2_PLAFLK	P41001 plasmodium
4	92.5	13.3	2468	1	MAPB_HUMAN	P46821 homo sapien
5	89	12.8	1888	1	YD12_SCHPO	Q14207 schizosacch
6	87.5	12.6	1702	1	IGA2_HAEIN	P45384 haemophilus
7	87	12.5	2464	1	MAPB_MOUSE	P14873 mus musculus
8	86	12.4	443	1	GLNA_PYRKO	Q08467 pyrococcus
9	85	12.2	720	1	IF2_STAEP	Q8C8t4 staphylococ
10	85	12.2	1202	1	DPOM_ASCIM	P23374 ascobolus i
11	84.5	12.1	1332	1	SP77 YEAST	P35177 saccharomyc
12	84.5	12.1	1694	1	IGA0_HAEIN	P44969 haemophilus
13	84	12.1	778	1	YFR8 YEAST	P43610 saccharomyc
14	83.5	12.0	479	1	U2R1 HUMAN	Q15695 homo sapien
15	83	11.9	258	1	IF31 HUMAN	O75822 homo sapien
16	83	11.9	270	1	TONE_HAEIN	P42872 haemophilus
17	82.5	11.9	439	1	GLNA_PYRAB	Q9ay99 pyrococcus
18	82	11.8	279	1	YMB4 YEAST	P49577 saccharomyc
19	82	11.8	427	1	YKT5 YEAST	P36046 saccharomyc
20	82	11.8	439	1	GLNA_PYRFU	Q05907 pyrococcus
21	82	11.8	439	1	GLNA_PYRWO	P36687 pyrococcus
22	82	11.8	949	1	IF2_HELPJ	Q9zm46 helicobacte
23	81.5	11.7	572	1	LMU1 HUMAN	P29536 homo sapien
24	81.5	11.7	752	1	DRS1 YEAST	P32892 saccharomyc
25	81.5	11.7	2459	1	MAPB RAT	P15205 rattus norv
26	81	11.6	797	1	PAT1 YEAST	P25644 saccharomyc
27	81	11.6	853	1	YCG1 YEAST	P25588 saccharomyc
28	81	11.6	1325	1	YAE6 SCHPO	Q09847 schizosacch
29	80.5	11.6	256	1	TONE_PASMU	Q9zb79 pasteurella
30	80.5	11.6	793	1	YF06_MYCPN	P75280 mycoplasma
31	80.5	11.6	914	1	PBPA_BACSU	P39793 bacillus su
32	80.5	11.6	1006	1	RAT1 YEAST	Q02792 saccharomyc
33	80	11.5	1688	1	LIP_STAEP	Q02510 staphylococ

1

78 KKDNPVNH SOLNESH RKE D LO REEHSOKSPSTKDVTATVLD--KNNISSKSTN--NP 132

```

Db 1150 SHDKRPHSKVE-----QKSSSEKSDNDKDLTHILDFVQNNFSSRIFPNKLLSP 1201
QY 133 NK 134
Db 1202 QK 1203

RESULT 2
CYL2 HUMAN
ID CYL2 HUMAN STANDARD; PRT; 348 AA.
AC Q14093;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cyclicin II (Multiple-band polypeptide II).
GN CYL2 OR CYL2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95255491; PubMed=7737358;
RA Hess H., Heid H., Zimbelmann R., Franke W.W.;
RT "The protein complexity of the cytoskeleton of bovine and human sperm
RT heads: the identification and characterization of cyclicin II.";
RL Exp. Cell Res. 218:174-182(1995).
CC -!- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
CC BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: CYLX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC -!- TISSUE SPECIFICITY: Testis.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; Z46788; CAA86752.1; -
CC PIR; I37271; I37271.
CC Genew; HGNC:2583; CYLC2.
CC MIM; 604035; -
CC GO; GO:0005200; F:Structural constituent of cytoskeleton; TAS.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT DOMAIN 25 347 31 X 3 AA REPEATS OF K-K-X.
FT DOMAIN 157 240 3 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 157 184 1.
FT REPEAT 185 212 2.
FT REPEAT 213 240 3.
SQ SEQUENCE 348 AA; 39079 MW; D86766599C1809E7 CRC64;

Query Match 13.6%; Score 95; DB 1; Length 348;
Best Local Similarity 30.9%; Pred. No. 1.6;
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 19 KGELEKGYQFGWEISGFEKKDAGYVNLKDTFKVPFKKIEEENKPTF---DV 75
Db 205 ESEGEKG-----GTEKDSKSAIELOAVRADEKDEKDKOANKGDE 256
QY 76 SK--KKNDPQVNHSLN-----ESHRKEDLQREHSQKSDTKD---VTATVLDKNNI 123
Db 257 SKDAKDAKEIKGKKDKKPSSTSDSKDDVKKE---SKDKATKDAKVAKOTEKESA 313
QY 124 SSK 126
Db 314 DSK 316

RESULT 3

```

```

TOP2 PLAFK
ID TOP2 PLAFK STANDARD; PRT; 1398 AA.
AC P41001;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94316496; PubMed=8041616;
RA Chiesman S., McAlleese S., Goman M., Johnson D., Horrocks P.,
RA Ridley R.G., Kilbey B.J.;
RL "The gene encoding topoisomerase II from Plasmodium falciparum.";
RL Nucleic Acids Res. 22:2547-2551(1994).
CC -!- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -!- SIMILARITY: Belongs to the type II topoisomerase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; X79345; -; NOT_ANNOTATED_CDS.
CC HSSP; P06786; 1BGW.
CC InterPro; IPR003594; ATPbind ATPase.
CC InterPro; IPR003957; CBFA_NFYB_topis.
CC InterPro; IPR001241; DNA_topoisoi.
CC InterPro; IPR002205; DNA_topoisoi.
CC Pfam; PF00204; DNA_gyraseb; 1.
CC Pfam; PF00521; DNA_topoisoi; 1.
CC Pfam; PF02518; HATPase_c; 1.
CC PRINTS; PR00615; CCAATSUBUNTA.
CC PRINTS; PR00418; TPI2FAMILY.
CC ProDom; PD000742; DNA_topoisoi; 1.
CC SMART; SM00387; HATPase_c; 1.
CC SMART; SM00433; TOP2c; 1.
CC SMART; SM00434; TOP4c; 1.
CC PROSITE; PS00177; TOPOISOMERASE II; 1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
FT NP_BIND 144 149 ATP (POTENTIAL).
FT ACT_SITE 830 830 DNA CLEAVAGE (BY SIMILARITY).
FT DOMAIN 271 281 POLY-ASN.
FT DOMAIN 308 316 POLY-ASN.
FT DOMAIN 1089 1093 POLY-LYS.
FT DOMAIN 1227 1234 POLY-LYS.
SQ SEQUENCE 1398 AA; 161029 MW; BAAD7BEE88FE5BE9 CRC64;

Query Match 13.4%; Score 93; DB 1; Length 1398;
Best Local Similarity 24.6%; Pred. No. 9.8;
Matches 35; Conservative 33; Mismatches 48; Indels 26; Gaps 6;

QY 8 VSEEDFI--PVYKGELEKGYQFGWEISGFEKKDAGYVNLKDTFKVPFKKIEE-- 63
Db 1142 VKDYDLLSMPIFSLTLEK---VEDLLTQLKEKEKELEILRNITVETWMLKDIKVEBAI 1198
QY 64 -----KKEENKPTFDVSKKNDPQVNHSLNESHKEDLQREHSQ---KSDSTK 111

```

Db 1199 EFQNVLSNREESNK--FKVARKQ-----GPSMKKKKKKKKLLSSDESEGGDTSDSSE 1251

Qy 112 DVATVLDKNNISSKSTNNPN 133

Db 1252 FLVNTLNIKKNTNKKYTTSSNN 1273

RESULT 4

MAPB HUMAN

ID MAPB\_HUMAN STANDARD; PRT; 2468 AA.

AC P46821; 1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain

DE LC1]

GN MAP1B.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal brain;

RX MEDLINE=95104835; PubMed=7806212;

RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;

RT "Cloning of human microtubule-associated protein 1B and the

RT identification of a related gene on chromosome 15.;"

RL Genomics 22:273-280 (1994).

CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.

CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES

CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST

CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS

CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN

CC STABILIZING MICROTUBULES.

CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE

CC WITH MAP1A AND MAP1B PROTEINS.

CC -!- DOMAIN: Has a highly basic region with many copies of the sequence

CC KKEE and KKEI/V, repeated but not at fixed intervals, which is

CC responsible for the binding of MAP1B to microtubules.

CC -!- PM: LC1 is coexpressed with MAP1B. It is a polypeptide generated

CC from MAP1B by proteolytic processing. It is free to associate with

CC both MAP1A and MAP1B. It interacts with the amino-terminal region

CC of MAP1B (By similarity).

CC -!- SIMILARITY: TO MAP1A.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL; L06237; AAA18904.1; .

CC Genew; HGNC:6836; MAP1B.

CC MIN; 157129; .

CC GO; GO:0005875; C:Microtubule associated complex; TAS.

CC InterPro; IPR000102; MAP1B neuraxin.

CC Pfam; PF00414; MAP1B neuraxin; 10.

CC PROSITE; PS00230; MAP1B NEURAXIN; 6.

CC Microtubules; Repeat; Phosphorylation.

CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.

REPEAT 1878 1894 MAP1B 1.

REPEAT 1895 1911 MAP1B 2.

REPEAT 1912 1928 MAP1B 3.

REPEAT 1929 1945 MAP1B 4.

REPEAT 1946 1962 MAP1B 5.

REPEAT 1963 1979 MAP1B 6.

REPEAT 1997 2013 MAP1B 7.

REPEAT 2014 2030 MAP1B 8.

REPEAT 2031 2047 MAP1B 9.

REPEAT 2048 2064 MAP1B 10.

FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY

FT KKEE AND KKEI/V REPEATS).

SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;

Query Match 13.3%; Score 92.5; DB 1; Length 2468;

Best Local Similarity 31.0%; Pred. No. 19;

Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

Qy 48 LSKDTFIKVPFKKIEEKEENKPTFDVSKKKONPQVNHSHKEDLQRE-----E 102

Db 638 VKKETVKP-----EDKKEKEKKEVAKEDKTPI---KKEKPKKEVKKEVKEIK 689

Qy 103 HSQKSSTKDV 113

Db 690 KEEKKPKKEV 700

RESULT 5

YDT2 SCHPO

ID YDT2 SCHPO STANDARD; PRT; 1888 AA.

AC O14207;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein C6B12.02c on chromosome 1.

GN SPAC6B12.02C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OX NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,

RA Sgourtos J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,

RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,

RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,

RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,

RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,

RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,

RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,

RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,

RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,

RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,

RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,

RA Cerrutti L., Lowe I., McCombie W.R., Paulsen I., Potashkin J.,

RA Sipakowski G.V., Ussery D., Barrel B.G., Nurse P.;

RT "The genome sequence of Schizosaccharomyces pombe.;"

RL Nature 415:871-880(2002).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

CC EMBL; Z98531; CAB11064.1; .

DR PIR; T39009; T39009.



```

CC -!- SIMILARITY: TO MAP1A.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X51396; CRA35761.1; -.
DR PIR; S07549; CRMSP1.
DR MGD; MGI:1306778; Mcap1b.
DR GO; GO:0016358; P:dendrite morphogenesis; IMP.
DR GO; GO:0001578; P:microtubule bundling; IMP.
DR InterPro; IPR001012; MAP1B neuraxin.
DR Pfam; PF00414; MAP1B neuraxin; 10.
DR PROSITE; PS00230; MAP1B NEURAXIN; 7.
DR Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2464 MAP1 LIGHT CHAIN LCL.
FT REPEAT 1874 1890 MAP1B 1.
FT REPEAT 1891 1907 MAP1B 2.
FT REPEAT 1908 1924 MAP1B 3.
FT REPEAT 1925 1941 MAP1B 4.
FT REPEAT 1942 1958 MAP1B 5.
FT REPEAT 1959 1975 MAP1B 6.
FT REPEAT 1993 2009 MAP1B 7.
FT REPEAT 2010 2026 MAP1B 8.
FT REPEAT 2027 2043 MAP1B 9.
FT REPEAT 2044 2060 MAP1B 10.
FT DOMAIN 589 787 KKEE AND KKEI/V REPEATS).
SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDBDA87 CRC64;

Query Match 12.5%; Score 87; DB 1; Length 2464;
Best Local Similarity 28.4%; Pred. No. 51;
Matches 29; Conservative 16; Mismatches 37; Indels 20; Gaps 4;

Qy 48 LSKDTFTKPVPK-KIEEKKEENKPTFDVSKKDKNPVNHSQLNESHKEDLORE----- 101
Db 632 VTOKVVKKEIKTKEEKKEE-KRPKEVVKKEKDTPL---KDEKPKKEVKKEIKKEI 686

Qy 102 -----EHSQKSDSTQVDTATVLDKNISSTKSTNNPNK 134
Db 687 KKEERKELKKEVKETPLKDAKKEVKKEEKEKKEPKK 728

RESULT 8
GLNA_PVRKO STANDARD; PRT; 443 AA.
AC O08467;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
GN GLNA.
OS Pyrococcus kodakaraensis.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Thermococcus.
OX NCBI_TaxID=69014;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOD1;
RX MEDLINE=97316461; PubMed=9172372;
RA Adul Rahman R.N.Z., Jongsareejit B., Fujiwara S., Imanaka T.;
RT "Characterization of recombinant glutamine synthetase from the
RT hyperthermophilic archaeon Pyrococcus sp. strain KOD1.";
RL Appl. Environ. Microbiol. 63:2472-2476(1997).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D86222; BAA20530.1; -.
DR HSP; P06201; ILGR.
DR InterPro; IPR001691; GLN synth.
DR InterPro; IPR004809; GlnA_adenyltn.
DR InterPro; IPR001637; GlnA_adenyltn.
DR Pfam; PF00120; gln-synt; 1.
DR Pfam; PF03951; gln-synt N; 1.
DR ProDom; PD001057; Gln_synt C; 1.
DR TIGRFAMs; TIGR00653; GlnA_1.
DR PROSITE; PS00180; GLNA_1; 1.
DR PROSITE; PS00181; GLNA_ATP; 1.
KW Ligase.
FT BINDING 362 362 AMP (UNDER CONDITIONS OF ABUNDANT
FT BINDING 443 AA; 50259 MW; 9426DCCFEF18168 CRC64;
SQ SEQUENCE 443 AA; 50259 MW; 9426DCCFEF18168 CRC64;

Query Match 12.4%; Score 86; DB 1; Length 443;
Best Local Similarity 17.8%; Pred. No. 10;
Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;

Qy 18 YKSELEKGYQFDGWEISGPEGKDKAGVINLSKDTFT----- 54
Db 42 YEEAEDGVSGFDGSSIPGFEIGEDSLIFKADPSTYAEIPWEGIGRVGYIKGDEPYQA 101

Qy 55 --KPVFKKIEKKKEEN-----KPTFDVSKKKD----- 80
Db 102 DPRGILKRVLEKEGLKAHIGPEFYFKNGTWELHIPSDDGGYFDLVGLDKAREIR 161

Qy 81 -----NPQVNHSQLNESHKEDLOREHSQKSD---STKDVATATVLD 119
Db 162 REIATLYMPYLGLKPEVLHHEVGKAQHEIDFRYDEALRTADNIVSFKHVRAVAE 215

RESULT 9
ID_IF2_STAEP STANDARD; PRT; 720 AA.
AC OBCST4;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-2.
GN INF2 OR SE0945.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016747; AA04542.1; -.
DR HAMAP; MF_00100; -.
DR PFam; PF00009; GTP_EFTU; 1.
DR PFam; PF03144; GTP_EFTU_D2; 1.
DR PFam; PF04760; IF2_N_2; 1.
DR ProDom; PD186100; IF2_N_1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 225 373 G-DOMAIN.
FT NP_BIND 231 238 GTP (BY SIMILARITY).
FT NP_BIND 277 281 GTP (BY SIMILARITY).
FT NP_BIND 331 334 GTP (BY SIMILARITY).
SQ SEQUENCE 720 AA; 79343 MW; 07FB5A6A59CF970C CRC64;

Query Match 12.2%; Score 85; DB 1; Length 720;
Best Local Similarity 28.0%; Pred. No. 20;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;

QY 61 IEEKKEENKPTFDVSKKD-----NPQVNHSQLNESHKEDLQREHSOKSDSTK----D 112
DB 38 LEEQIKALDKKFKASQAKDTNKQNTQNNHQSKNNKQNSDKKEKQSKNNKPTKKKEQN 97
QY 113 VTATVLDKNNISKSSTNNPNK 134
DB 98 NKGKQKNNKNNKNNKNNKNNK 119

RESULT 10
DPOM_ASCIM
ID DPOM_ASCIM STANDARD; PRT; 1202 AA.
AC P22374;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Probable DNA polymerase (EC 2.7.7.7).
OS Ascobolus immersus.
OG Mitochondrion.
OG Plasmid pA12.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Pezizomycetes;
OC Pezizales; Ascobolaceae; Ascobolus.
OX NCBI_TaxID=5191;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2/1;
RX MEDLINE=90066356; PubMed=2573821;
RA Kempken F., Meinhardt F., Esser K.;
RT "In organello replication and viral affinity of linear,
RT extrachromosomal DNA of the ascomycete Ascobolus immersus.";
RL Mol. Gen. Genet. 218:523-530 (1989).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).
CC -!- MISCELLANEOUS: THIS DNA POLYMERASE REQUIRES A PROTEIN AS A PRIMER
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. STRONG. TO
CC DNA POLYMERASE OF OTHER FUNGAL AND PLANTS MITOCHONDRIAL PLASMIDS.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15982; CAA34106.1; -.
DR PIR; S05362; S05362.
DR InterPro; IPR006172; DNA_pol_B.
DR InterPro; IPR004868; DNA_pol_B_2.
DR -----

DR PFam; PF03175; DNA_pol_B_2; 1.
DR SMART; SM00486; POLBc; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Plasmid; Mitochondrion.
SQ SEQUENCE 1202 AA; 138279 MW; 51D41FCEBDBF2CDE CRC64;

Query Match 12.2%; Score 85; DB 1; Length 1202;
Best Local Similarity 25.6%; Pred. No. 34;
Matches 33; Conservative 21; Mismatches 55; Indels 20; Gaps 5;

QY 19 KGELEKGYQFDGWEISGPEGKKGADGYVNLNLSKDFIKPVPFKKIE-----EKKEENK-P 71
DB 372 KGSU-----SDFPKPLKTIETGTYANYTFPIKDIVKDKINKINPGLDLPKTMDLSKWP 427
QY 72 TFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDKNNI-----S 124
DB 428 NLKLNKDKTSGEIRMTIKNNQSYDI--IGHMIINDGENVITFNRAVDNSIIKIFTVTD 485

QY 125 SKSTNNPN 133
DB 486 SMGNTNDPN 494

RESULT 11
SPT7_YEAST
ID SPT7_YEAST STANDARD; PRT; 1332 AA.
AC P35177;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Transcriptional activator SPT7.
GN SPT7 OR YBR081C OR YBR0739.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95229044; PubMed=7713415;
RA Gansheroff L.J., Dollard C., Tan P., Winston F.;
RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein
RT important for transcription in vivo.";
RL Genetics 139:523-536 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95076715; PubMed=7985423;
RA van der Aart Q.J.M., Barthe C., Doignon F., Aigle M., Crouzet M.,
RA Steensma H.Y.;
RT "Sequence analysis of a 31 kb DNA fragment from the right arm of
RT Saccharomyces cerevisiae chromosome II.";
RL Yeast 10:959-964 (1994).
RN [3]
RP SEQUENCE OF 1-835 FROM N.A.
RC STRAIN=S288C;
RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
RA Vissers S.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 463-523 FROM N.A.
RX MEDLINE=92285152; PubMed=1350857;
RA Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J.,
RA David I.B.;
RT "The bromodomain: a conserved sequence found in human, Drosophila and
RT yeast proteins.";
RL Nucleic Acids Res. 20:2603-2603 (1992).
CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY
CC OTHER GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -----
```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; L22537; AAC37424.1; -;  
 DR EMBL; X76294; CRA53940.1; -;  
 DR EMBL; Z35950; CRA85026.1; -;  
 DR EMBL; M87651; AAA35087.1; -;  
 DR PIR; S41552; S41552.  
 DR HSSP; Q92831; 1B91.  
 DR TRANSFAC; T04835; -;  
 DR SGD; S0000285; SPT7.  
 DR GO; GO:0000124; C:SAGA complex; IDA.  
 DR InterPro; IPR001487; Bromodomain.  
 DR Pfam; PF00439; bromodomain; 1.  
 DR PRINTS; PR00503; BROMODOMAIN.  
 DR SMART; SM00297; BROMO; 1.  
 DR PROSITE; PS00633; BROMODOMAIN 1; 1.  
 DR PROSITE; PS0014; BROMODOMAIN 2; 1.  
 KW Transcription regulation; Nuclear protein; Activator; Bromodomain.  
 FT DOMAIN 458 528 BROMODOMAIN.  
 SQ SEQUENCE 1332 AA; 152616 MW; 083B63624669244F CRC64;

Query Match 12.1%; Score 84.5; DB 1; Length 1332;  
 Best Local Similarity 24.3%; Pred. No. 41;  
 Matches 36; Conservative 24; Mismatches 61; Indels 27; Gaps 6;

QY 6 TIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKKDAGYVNLKSDTFIK---PVFKKIE 62  
 DB 550 TIRRADLEKIEDEMEKDKDYELDEEERVAGSRKG-----LNGAHMLAKENGKVKSEKDS 605  
 QY 63 EKKEEENKPTFD-----VSKKKD-----NPQVNHSQLNESHKEDLQR-EEH 103  
 DB 606 SKTVKDEAPTNDKLTSTVPEGEKEKDTASSTVTVHENVNKNKENGKNEEQDMVEES 665  
 QY 104 SKQSDSTKDVATVLD-KNNLSKSTTN 130  
 DB 666 SKTEDSSKDAADAKKDTDLGLQDKTAEN 693

RESULT 12  
 ID IGAO HAEIN STANDARD; PRT; 1694 AA.  
 AC P44969;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAI protease).  
 GN IGAI OR IGAI OR HI0990.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OK NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Serotype D;  
 RA Wright A., Fishman Y., Tai F., Plaut A.G.;  
 RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.  
 RP [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA McLavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.P., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.;

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512 (1995).  
 CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A  
 CC PRODUCING INTACT FC AND FAB FRAGMENTS.  
 CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
 CC certain Pro-Xaa bonds in the hinge region. No small molecule  
 CC substrates are known.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
 CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
 CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
 CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X59800; -; NOT ANNOTATED\_CDS.  
 DR EMBL; U32779; AAC22651.1; -;  
 DR PIR; H64106; H64106.  
 DR MEROPS; S06.001; -;  
 DR TIGR; HI0990; -;

DR InterPro; IPR006315; Autotransport.  
 DR InterPro; IPR005546; Autotransporter.  
 DR InterPro; IPR000710; IGA-S6.  
 DR InterPro; IPR004899; Pertactin.  
 DR Pfam; PF03797; Autotransporter; 1.  
 DR Pfam; PF02395; IGAI; 1.  
 DR Pfam; PF03212; Pertactin; 1.  
 DR PRINTS; PR00321; IGASERPTASE.  
 DR TIGRFAMs; TIGR01414; autotrans\_barl; 1.  
 DR Hydroxylase; Serine protease; transmembrane; Zymogen; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.  
 FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).  
 FT ACT\_SITE 288 288 PROBABLE.  
 FT CONFLICT 253 254 EN -> GV (IN REF. 1).  
 FT CONFLICT 272 272 G -> A (IN REF. 1).  
 FT CONFLICT 464 464 G -> E (IN REF. 1).  
 FT CONFLICT 866 866 S -> T (IN REF. 1).  
 FT CONFLICT 1036 1036 A -> D (IN REF. 1).  
 FT CONFLICT 1074 1074 A -> G (IN REF. 1).  
 FT CONFLICT 1421 1421 A -> G (IN REF. 1).  
 FT CONFLICT 1545 1545 H -> T (IN REF. 1).  
 SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;

Query Match 12.1%; Score 84.5; DB 1; Length 1694;  
 Best Local Similarity 26.1%; Pred. No. 53;  
 Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;

QY 46 INLSKDTFKIPVFKKIEKKEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHS- 104  
 DB 1288 INTGSAITETAEKSKDPQTETAASTDASQHKANTVADNSVANNSESSDPKRRRSI 1347

QY 105 --QKSDSTKDVATVLDKNNISSKSTTNPNK 134

DB 1348 SQPQTSABETTAASTDDETTIADNSKSKPNR 1379

RESULT 13

YFK8 YEAST

ID YFK8 YEAST STANDARD; PRT; 778 AA.

AC P43610;

```

RESULT 14
U2R1 HUMAN STANDARD; PRT; 479 AA.
ID U2R1 HUMAN
AC Q15695; Q13570;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit
DE related-protein.1.
DE UZAFI-RS1 OR UZAFIRS1 OR U2AFBPL.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP TISSUE=Brain;
RC MEDLINE=96163878; PubMed=8586425;
RA Kitagawa K., Wang X., Hatada I., Yamaoka T., Nojima H.,
RA Inazawa J., Abe T., Mitsuya K., Oshimura M., Murata A., Monden M.,
RA Mukai T.;
RT "Isolation and mapping of human homologues of an imprinted mouse gene
RT Uzaafi-rsl.";
RL Genomics 30:257-263(1995).
[2]
SEQUENCE FROM N.A.
MEDLINE=96212931; PubMed=8630064;
RA Pearall R.S., Shibata H., Brozowska A., Yoshino K., Okuda K.,
RA Dejong P.J., Plass C., Chapman V.M., Hayashizaki Y., Held W.A.;
RT "Absence of imprinting in U2AFBPL, a human homologue of the imprinted
RT mouse gene U2afbp-rs.";
RL Biochem. Biophys. Res. Commun. 222:171-177(1996).
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -1- SIMILARITY: Contains 2 C3H1-type zinc fingers.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement.(see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D49676; BAA08532.1; -.
DR EMBL; U51224; AAA98669.1; -.
DR MIM; 601079; -.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003723; F:RNA binding activity; NAS.
DR InterPro; IPR000504; RNA_rec_mot.
DR InterPro; IPR000571; Znf_CCCH.
DR Pfam; PF00076; rrm; 1.
DR Pfam; PF00642; zf-CCCH; 2.
DR SMART; SM00360; RRM; 1.
DR SMART; SM00356; ZnF_C3H1; 2.
DR PROSITE; PS00102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; FALSE NEG.
DR Nuclear protein; RNA-BINDING; Ribonucleoprotein; Zinc-finger; Repeat.
FT DOMAIN 203 309 RNA-BINDING (RRM).
SO SEQUENCE 479 AA; 57643 MW; 96F326694B04E7C0 CRC64;

```

[illegible]

```

DR ENBL; BC002719; AA02719.1; -.
DR DR Genes; HGNC:3270; EIF3S1.
DR GK; 075822; -.
DR MTM; 603910; -.
DR DR GO; GO:0005852; C:eukaryotic translation initiation factor 3 . . .; TAS.
DR DR GO; GO:0003743; F:translation initiation factor activity; TAS.
DR DR GO; GO:0006446; P:regulation of translational initiation; TAS.
DR KW Initiation factor; Protein biosynthesis.
FT DOMAIN 2 8 POLY-ALA.
FT FT DOMAIN 29 32 POLY-GLY.
FT FT DOMAIN 53 57 POLY-ASP.
FT FT DOMAIN 218 224 POLY-LYS.
FT FT CONFLICT 40 40 E -> G (IN REF. 1).
FT FT CONFLICT 141 141 A -> T (IN REF. 2).
SQ SEQUENCE 258 AA; 29062 MW; 8362423542445AA CRC64;

Query Match 11.9%; Score 83; DB 1; Length 258;
Best Local Similarity 25.4%; Pred. No. 9.6;
Matches 30; Conservative 25; Mismatches 51; Indels 12; Gaps 4;

QY 18 YKGELEKGYQFGWETSGFEGKKDAGY--VINLSKDTFTKVPVKIEEKKEENKPTFDV 75
Db 39 WEGEDEDEVDKNWDDDDDEKKEAEVKEVISEK---KKAIEKIEKKEKQKQKQERI 95
QY 76 SKKKNPQ-----VNHSQLNESHKEDIQREHRSQKSDTKDVTATV--LDKNINISK 126
Db 96 KRLSEPEEPKVLTPPEQLAKLRLKKLQFESLELAKETFGVNNVAYGIDAWNPSR 153

RESULT 16
TOMB_HAEBIN STANDARD; PRT; 270 AA.
ID TOMB_HAEBIN STANDARD; PRT; 270 AA.
AC P42872;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE TonB protein.
DE TOMB OR HI0251.
GN Haemophilus influenzae.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NTHI TN106;
RX MEDLINE=94245357; PubMed=8188372;
RA Jarosik G.P., Sanders J.D., Cope L.D., Muller-Eberhard U.,
RA Hansen E.J.;
RT "A functional tonB gene is required for both utilization of heme and
RT virulence expression by Haemophilus influenzae type b.";
RL Infect. Immun. 62:2470-2477(1994).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Karlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER

```

```
CC MEMBRANE PROTEINS. REQUIRED FOR HEME UTILIZATION AND VIRULENCE.
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC
CC PERIPLASM VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
CC PERIPLASM (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TONB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U04996; AAA60460.1; -.
CC DR EMBL; U32711; AAC21917.1; -.
CC DR TIGR; H10251; -.
CC
CC InterPro: IPR003538; TonB.
CC DR PFam; PF03544; TonB; 1.
CC DR PRINTS; PR01374; TONBPROTEIN.
CC DR TIGRFAMs; TIGR01352; tonB Cterm; 1.
CC KW Transport; Protein transport; Inner membrane; Periplasmic;
CC Transmembrane; Signal-anchor; Repeat; Virulence; Complete proteome.
CC FT DOMAIN 1 7 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 8 27 SIGNAL-ANCHOR (POTENTIAL).
CC FT DOMAIN 28 270 PERIPLASMIC (POTENTIAL).
CC FT DOMAIN 70 77 4 X 2 AA TANDEM REPEATS OF E-P.
CC FT DOMAIN 58 90 GLU/PRO-RICH.
CC FT DOMAIN 98 143 GLU/LYS/PRO-RICH.
CC FT VARIANT 17 17 A -> V (IN STRAIN TN106).
CC FT VARIANT 31 31 E -> K (IN STRAIN TN106).
CC FT VARIANT 66 66 D -> N (IN STRAIN TN106).
CC FT VARIANT 76 84 MISSING (IN STRAIN TN106).
CC FT VARIANT 116 116 G -> E (IN STRAIN TN106).
CC FT VARIANT 120 120 G -> E (IN STRAIN TN106).
CC FT VARIANT 134 134 K -> Q (IN STRAIN TN106).
CC FT VARIANT 140 140 E -> D (IN STRAIN TN106).
CC FT VARIANT 146 146 E -> K (IN STRAIN TN106).
CC FT VARIANT 226 226 K -> R (IN STRAIN TN106).
CC SQ SEQUENCE 270 AA; 29193 MW; 4CDA54B46F7D10A0 CRC64;
CC
CC Query Match 11.9%; Score 83; DB 1; Length 270;
CC Best Local Similarity 32.1%; Pred. No. 10;
CC Matches 26; Conservative 16; Mismatches 33; Indels 6; Gaps 3;
CC
CC QY 50 KDTFTKVPKKEEKEENKPTFDVSKKKNPOVNHSQLNESHKEDLQREHSQKSDS 109
CC DB 93 EDPTTKPEPKKEPEKEPKPK---GKPKGPK-NKPKVKPKPKINKE--LPKGD 146
CC
CC QY 110 TKDVTATVLDKNNISSKSTTN 130
CC DB 147 NIDSSANVDKASTTSAANSN 167
CC
CC RESULT 17
CC GLNA PYRAB STANDARD; PRT; 439 AA.
CC ID GLNA PYRAB STANDARD; PRT; 439 AA.
CC AC Q9UY59;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
CC DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
CC GN GLNA OR PYRAB16090 OR PAB1292.
CC OS Pyrococcus abyssi.
CC OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
CC PYROCOCUS.
CC OX NCBI_TaxID=29292;
CC [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=GE5 / Orsay;
CC RA PubMed=12622808;
```

```
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ248288; CAB50513.1; -.
CC DR PIR; C75009; C75009.
CC DR HSP; P06201; ILGR.
CC DR InterPro; IPR001691; GLN synth.
CC DR InterPro; IPR004809; GlnA.
CC DR InterPro; IPR001637; GlnA adenyltn.
CC DR PFam; PF00120; gln-synt; 1.
CC DR PFam; PF03951; gln-synt_N; 1.
CC DR ProDom; PD001057; Gln_synt_C; 1.
CC DR TIGRFAMs; TIGR00653; GlnA; 1.
CC DR PROSITE; PS00180; GLNA_1; 1.
CC DR PROSITE; PS00181; GLNA_ATP; 1.
CC KW Ligase; Complete proteome.
CC FT BINDING 358 358 AMP (UNDER CONDITIONS OF ABUNDANT
CC BINDING 358 358 GLUTAMINE) (BY SIMILARITY).
CC SQ SEQUENCE 439 AA; 49742 MW; 62CCFD3970A98AF0 CRC64;
CC
CC Query Match 11.9%; Score 82.5; DB 1; Length 439;
CC Best Local Similarity 17.8%; Pred. No. 18;
CC Matches 28; Conservative 24; Mismatches 36; Indels 69; Gaps 4;
CC
CC QY 22 LEKGYPQFGWEISGPEKKDAGYVINLSKDTFI-----KP----- 56
CC DB 42 IEDGISFDGSSVPGFGIEDSLIFKADPDYVEVPDNNVARYGYIKDGKPYCADPRG 101
CC
CC QY 57 VFKEIEKKEEN-----KPTFDVSKKKN----- 81
CC DB 102 VLKRVIEKLAEMGKAYIGPEPEFYLFKNGSWELEIPDVGGYFDILTLDKAKDIKREIA 161
CC
CC QY 82 -----POVNHSQLNESHKEDLQREHSQKSDS 109
CC DB 162 EYMPFGLVPELVHHEVGKAQHEIDFRYDEALKTADN 198
CC
CC RESULT 18
CC YME4 YEAST STANDARD; PRT; 279 AA.
CC ID YME4 YEAST STANDARD; PRT; 279 AA.
CC AC P49957;
CC DT 01-OCT-1996 (Rel. 34, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
CC DE Hypothetical 32.4 kDa protein in TAF40-ERV25 intergenic region.
CC GN YML014W OR YN9571.04.
CC OS Saccharomyces cerevisiae (Baker's yeast).
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
CC OX NCBI_TaxID=4932;
CC [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=S288c / AB972;
CC RX PubMed=9169872;
CC RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
CC Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
CC Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
```



Query Match 11.8%; Score 82; DB 1; Length 439;  
 Best Local Similarity 18.1%; Pred. No. 20;  
 Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

QY 25 GYQFDGWEISGEGKKGAGYVNLKDTFI-----KP-----VFK 59  
 DB 45 GISFGSSVPGFQIGSDSLVFKADPDYVEVPDWNVARVGYFYKDKPYGADPRGILK 104  
 QY 60 KIEKKEBEN-----KPTFDVSKKKDN-----81  
 DB 105 RALELEKEGKYKAYIGPEPEFYLFKNGTWLEIPDVGGYFDILTLDKARDIRREIAEYM 164  
 QY 82 -----POVNHSQLNESHRKEDLOREHSQKSD---STKQVTVATVLDKNNI 123  
 DB 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKTADNIVSFKYITKAVAEHGL 215

RESULT 21  
 GLNA PYRMO STANDARD; PRT; 439 AA.  
 ID GLNA PYRMO STANDARD; PRT; 439 AA.  
 AC P36687;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glutamine synthetase I (EC 6.3.1.2) (Glutamate--ammonia ligase) (GSI).  
 GN GLNA.  
 OS Pyrococcus woesei.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus  
 OX NCBI\_TaxID=2262;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93259940; PubMed=8098326;  
 RA Tiboni O., Cammarano P., Sanangelantoni A.M.;  
 RT "Cloning and sequencing of the gene encoding glutamine synthetase I  
 from the archaeum Pyrococcus woesei: anomalous phylogenies inferred  
 from analysis of archaeal and bacterial glutamine synthetase I  
 sequences.";  
 RL J. Bacteriol. 175:2961-2969 (1993).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
 L-glutamine.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 EMBL; X60161; CA442730.1; --  
 DR PIR; A36911; A36911.  
 DR HSSP; P06201; ILGR.  
 DR InterPro; IPR001691; GLN synth.  
 DR InterPro; IPR004809; GLN.  
 DR InterPro; IPR001637; GLN adenyln.  
 DR Pfam; PF00120; Gln-synt; 1.  
 DR Pfam; PF03951; Gln-synt; N; 1.  
 DR ProDom; PD001057; Gln synt; C; 1.  
 DR TIGRFAMs; TIGR00653; GLN; 1.  
 DR PROSITE; PS00180; GLNA.1; 1.  
 DR PROSITE; PS00181; GLNA\_ATP; 1.  
 KW Ligase.  
 FT BINDING 358 358 AMP (UNDER CONDITIONS OF ABUNDANT  
 GLUTAMINE). (BY SIMILARITY).  
 SQ SEQUENCE 439 AA; 50066 MW; C503B43B790BD25C CRC64;

Query Match 11.8%; Score 82; DB 1; Length 439;  
 Best Local Similarity 18.1%; Pred. No. 20;  
 Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

QY 25 GYQFDGWEISGEGKKGAGYVNLKDTFI-----KP-----VFK 59  
 DB 45 GISFGSSVPGFQIGSDSLVFKADPDYVEVPDWNVARVGYFYKDKPYGADPRGILK 104  
 QY 60 KIEKKEBEN-----KPTFDVSKKKDN-----81  
 DB 105 RALELEKEGKYKAYIGPEPEFYLFKNGTWLEIPDVGGYFDILTLDKARDIRREIAEYM 164  
 QY 82 -----POVNHSQLNESHRKEDLOREHSQKSD---STKQVTVATVLDKNNI 123  
 DB 165 PSFGLIPEVLHHEVGKAQHEIDFRYDEALKTADNIVSFKYITKAVAEHGL 215

RESULT 22  
 IF2\_HELPF STANDARD; PRT; 949 AA.  
 ID IF2\_HELPF STANDARD; PRT; 949 AA.  
 AC Q9ZM46;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Translation initiation factor IF-2.  
 GN INF2 OR JHP0377.  
 OS Helicobacter pylori J99 (Campylobacter pylori J99).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=85963;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99120557; PubMed=9923682;  
 RA Alm R.A., Ling L.-S.L., Meir D.T., King B.L., Brown E.D., Doig P.C.,  
 RA Smith D.R., Noonan B., Guld B.C., deJonge B.L., Carmel G.,  
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
 RA Trust T.J.;  
 RT "Genomic sequence comparison of two unrelated isolates of the human  
 gastric pathogen Helicobacter pylori.";  
 RL Nature 397:176-180 (1999).  
 CC -!- FUNCTION: One of the essential components for the initiation of  
 protein synthesis. Protects formylmethionyl-tRNA from spontaneous  
 hydrolysis and promotes its binding to the 30S ribosomal subunits.  
 CC Also involved in the hydrolysis of GTP during the formation of the  
 70S ribosomal complex (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 EMBL; AE001472; AAO05948.1; --  
 DR PIR; E71940; E71940.  
 DR HAMAP; MF\_00100; --; 1.  
 DR InterPro; IPR000795; EF GTPbind.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR InterPro; IPR000178; IF2\_N.  
 DR InterPro; IPR006847; IF2\_N.  
 DR InterPro; IPR005225; Small GTP.  
 DR Pfam; PF00009; GTP EFTU; 1.  
 DR Pfam; PF03144; GTP EFTU D2; 2.  
 DR Pfam; PF04760; IF2\_N; 1.  
 DR ProDom; PD186100; IF2; 1.  
 DR TIGRFAMs; TIGR00487; IF-2; 1.  
 DR TIGRFAMs; TIGR00231; small GTP; 1.  
 DR PROSITE; PS01176; IF2; 1.  
 KW Initiation factor; Protein biosynthesis; GTP-binding;  
 FT DOMAIN 451 599 G-DOMAIN.  
 FT NP\_BIND 457 464 GTP (BY SIMILARITY).

```
PT NP BIND 503 507 GTP (BY SIMILARITY).
FT NP BIND 557 560 GTP (BY SIMILARITY).
FT DOMAIN 321 326 POLY-GLU.
SQ SEQUENCE 949 AA; 105961 MW; FA8969B0C64B3278 CRC64;

Query Match 11.8%; Score 82; DB 1; Length 949;
Best Local Similarity 28.7%; Pred. No. 45;
Matches 29; Conservative 20; Mismatches 20; Indels 32; Gaps 5;

Qy 50 KDTPIKVFVKIEBKKEENKPTFDVSKKD-----NPO-VNHSQLES 92
Db 98 EETVAQPKPKTEKKKEAP-----IKKKEIIVTFNQTPLVNTPRVSHSQIEKA 154
Qy 93 HRKEDLQREHSQSDSTKDVATVLDKNINSSKSTNNPN 133
Db 155 --KQKLEIQKSREA-----LNKLTQSNTNTNNAN 183

RESULT 23
LMD1 HUMAN
ID LMD1 HUMAN STANDARD; PRT; 572 AA.
AC P29536;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Leiomodin 1 (Leiomodin, muscle form) (64 kDa autoantigen D1) (64 kDa
DE autoantigen ID) (64 kDa autoantigen ID3) (Thyroid-associated
DE ophthalmopathy autoantigen) (Smooth muscle leiomodin) (SM-Lmod).
GN LMOD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=91225220; PubMed=2026759;
RA Dong Q., Ludgate M., Vassart G.;
RT "Cloning and sequencing of a novel 64-kDa autoantigen recognized by
RT patients with autoimmune thyroid disease.";
RL J. Clin. Endocrinol. Metab. 72:1375-1381(1991).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=99451105; PubMed=10520227;
RA Conley C.A., Fowler V.M.;
RT "Localization of the human 64kD autoantigen D1 to myofibrils in a
RT subset of extraocular muscle fibers.";
RL Curr. Eye Res. 19:313-322(1999).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=21248187; PubMed=11350761;
RA Conley C.A.;
RT "Leiomodin and tropomodulin in smooth muscle.";
RL Am. J. Physiol. 280:C1645-C1656(2001).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=21218919; PubMed=11318603;
RA Conley C.A., Fritz-Six K.L., Almenar-Queralt A., Fowler V.M.;
RT "Leiomodins: larger members of the tropomodulin (Tmod) gene family.";
RL Genomics 73:127-139(2001).
CC -1- SUBCELLULAR LOCATION: CYTOSKELETON.
CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE (HEART, SKELETAL MUSCLE, COLON
CC AND SMALL INTESTINE), A SUBSET OF STRIATED MUSCLE FIBERS, AND AT
CC LOW LEVEL IN THYROID.
CC -1- DISEASE: RECOGNIZED BY PATIENTS WITH AUTOIMMUNE THYROID DISEASE.
CC -1- SIMILARITY: BELONGS TO THE TROPOMODULIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch.
CC EMBL; X54162; CAA38101.1; -.
DR PIR; S18732; S18732.
DR Genew; HGNC:6647; LMOD1.
DR MIM; 602715; -.
DR GO; GO:0005624; C.membrane fraction; TAS.
DR GO; GO:0009405; P.pathogenesis; TAS.
DR InterPro; IPR004934; Tropomodulin.
DR InterPro; IPR003124; WH2.
DR Pfam; PF03250; Tropomodulin; 1.
DR Pfam; PF02205; WH2; 1.
DR SMART; SM00246; WH2; 1.
KW Antigen; Repeat; Cytoskeleton.
FT DOMAIN 137 265 8 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 137 152 1.
FT REPEAT 153 168 2.
FT REPEAT 169 184 3.
FT REPEAT 185 200 4.
FT REPEAT 201 216 5.
FT REPEAT 217 232 6.
FT REPEAT 233 248 7.
FT REPEAT 249 265 8.
FT DOMAIN 480 499 5 X 4 AA APPROXIMATE TANDEM REPEATS.
SQ SEQUENCE 572 AA; 63737 MW; D4B42F880523DE94 CRC64;

Query Match 11.7%; Score 81.5; DB 1; Length 572;
Best Local Similarity 22.8%; Pred. No. 29;
Matches 38; Conservative 21; Mismatches 57; Indels 51; Gaps 6;

Qy 19 KGELEKGYQFDGWEISGFEK--KDAGYVINLSKDTFIKPVFKK-----IEEK 64
Db 99 RGGKKKFSRDRDEAGGSGEKPKEKIIRGIDKGRVRAAVDVKKEAGKDGGRGERAVATK 158
Qy 65 KEENK-----PTFVSKKKONPQVNHLSQNSHRKE----- 96
Db 159 KEBEKGGDRNTGLSRDKDKKREEMKEVAKDEKVKERRNTDTRKEGEMKMRAGNT 218
Qy 97 DLQREHSQK-----SDSTKDVAT-----VLDKNINSSKSTNNPNK 134
Db 219 DMKKEDEKVRKGTNTDTRKDKDEKVKNEPLHEKEAKDSKTPTPEK 265

RESULT 24
DRS1 YEAST
ID DRS1 YEAST STANDARD; PRT; 752 AA.
AC P32892;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable ATP-dependent RNA helicase DRS1.
GN DRS1 OR YLL008W OR L1345.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93087480; PubMed=1454790;
RA Ripmaster T.L., Vaughn G.P., Woolford J.L. Jr.;
RT "A putative ATP-dependent RNA helicase involved in Saccharomycetes
RT cerevisiae ribosome assembly.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:11131-11135(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY23;
RX MEDLINE=96405918; PubMed=8810043;
RA Miosga T., Zimmermann F.K.;
RT "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
RT a 43.7 kb fragment of chromosome XII including an open reading frame
RT homologous to the human cystic fibrosis transmembrane conductance
RT regulator protein CFTR.";
RL Yeast 12:693-708(1996).
```



```

RN [3]
RC SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RA MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albertmann K., Andre B., Ansgorge W.,
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Etian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Nentwich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M.,
RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wamburt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hoheisel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
RL Nature 387:87-90(1997).
CC -!- FUNCTION: PROBABLE HELICASE INVOLVED IN RIBOSOME ASSEMBLY.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L00683; AAA34666.1; -
CC EMBL; X91488; CAA62783.1; -
CC EMBL; 273113; CAA97452.1; -
CC PIR; S64750; S64750.
CC HSP; Q58083; 1HV8.
CC SGD; S0003931; DRS1.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR000629; DEAD_box.
CC InterPro; IPR001650; Helicase_C.
CC Pfam; PF00270; DEAD; 1.
CC Pfam; PF00271; helicase_C; 1.
CC SMART; SM00487; HELIC; 1.
CC SMART; SM00490; HELIC; 1.
CC PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
KW ATP-binding, RNA-binding, Helicase, Nuclear protein.
FT DOMAIN 170 190 POLY-GLU.
FT NP_BIND 275 282 ATP (POTENTIAL).
FT SITE 385 388 DEAD BOX.
FT CONFLICT 1 68 MVVGTKKYKSNLDFVPTIGSDSDVPLDSSDDEKVEAKTT
FT KKRKGKNNKKYVSEGDNLDEDVHDLDD -> MTKKSLRLRL
FT RRSGRVRIIRKRLVRGITSMMRFRWT (IN REF. 1).
SQ SEQUENCE 752 AA; 84843 MW; 60747607A6E5E48 CRC64;
Query Match 11.7%; Score 81.5; DB 1; Length 752;
Best Local Similarity 21.0%; Pred. No. 38;
Matches 30; Conservative 26; Mismatches 48; Indels 39; Gaps 6;
QY 8 VSEDFILPVYKGELEKGVQFD-----GWEI--SGPEGKKDAGYVNLKSTFTI 54
Db ||| : : : : : ||| : : : : : ||| : : : : :
53 VSEGNLDVHEDLDAGFKFDLADDTTNSFQGNFLAEGESNKDDAEAFV--KKQVDL 110
QY 55 KPVFKK-----IEKKEENKPTDVSKKDN-----PQVNHSQLNE 91
Db : : : : : : : : : : : : : : : : : : : : : : : : : : :
111 DKTIIRKGLVKMAHIDSKQEETKE-KVEKENDSDDEELAMDGFGMAPNNGDENOS 169
QY 92 SHRKEDLOREHESQKSDSTKDTV 114
Db : : : : : : : : : : : : : : : : : : : : : : : : : : :
170 EEEEEEEKKEEEEEEEQEMT 192
RESULT 25
MAPB RAT STANDARD; Q9ER21; Q9QW92; PRT; 2459 AA.
AC P15205; Q62958; Q9ER21; Q9QW92; PRT; 2459 AA.

```

```

DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP1B) (Neuraxin) [Contains: MAP1
DE light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_Taxid=10116;
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=96257242; PubMed=8666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
RT associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).
RN [2]
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Glial tumor;
RX MEDLINE=92347374; PubMed=1639092;
RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
RT recombinant rat MAP 1B.";
RL Eur. J. Cell Biol. 57:66-74(1992).
RN [3]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=Spinal cord;
RX MEDLINE=90059871; PubMed=2555150;
RA Rienitz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
RA Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Neuraxin, a novel putative structural protein of the rat central
RT nervous system that is immunologically related to microtubule-
RT associated protein 5.";
RL EMBO J. 8:2879-2888(1989).
RN [4]
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX MEDLINE=97405699; PubMed=9260743;
RA Ma D., Nothias F., Boyne L.J., Fischer I.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
RT in rat CNS and PNS during development.";
RL J. Neurosci. Res. 49:319-332(1997).
CC -!- FUNCTION: The function of brain MAPs is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC heart or muscle.
CC -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC nerve levels are high early in development but decrease during
CC postnatal development and are low in adults. In dorsal root
CC ganglia levels remain high throughout development.
CC -!- INDUCTION: By nerve growth factor.
CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (By similarity).
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: TO MAP1A.
CC -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC 2459) was originally described as neuraxin in Ref.3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```



the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```
CC EMBL; U52950; AAB17068.1; -.
CC EMBL; X60370; CAC16162.1; -.
CC EMBL; X16623; CAA34620.1; ALT_SEQ.
CC PIR; A56577; A56577.
CC InterPro; IPR000102; MAP1B neuraxin.
CC Pfam; PF00414; MAP1B neuraxin; 10.
CC PROSITE; PS00230; MAP1B NEURAXIN; 8.
CC Microtubules; Repeat; Phosphorylation.
CC CHAIN ? 2459 MAP1 LIGHT CHAIN LCL.
FT REPEAT 1869 1885 MAP1B 1.
FT REPEAT 1886 1902 MAP1B 2.
FT REPEAT 1903 1919 MAP1B 3.
FT REPEAT 1920 1936 MAP1B 4.
FT REPEAT 1937 1953 MAP1B 5.
FT REPEAT 1954 1970 MAP1B 6.
FT REPEAT 1988 2004 MAP1B 7.
FT REPEAT 2005 2021 MAP1B 8.
FT REPEAT 2022 2038 MAP1B 9.
FT REPEAT 2039 2055 MAP1B 10.
FT DOMAIN 559 1035 GLU-RICH.
FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
  KEE AND KEEI/V REPEATS).
FT DOMAIN 2224 2312 LYS-RICH.
FT CONFLICT 127 127 M -> V (IN REF. 1).
FT CONFLICT 140 140 R -> S (IN REF. 1).
FT CONFLICT 2112 2112 R -> K (IN REF. 3).
FT CONFLICT 2169 2169 L -> I (IN REF. 3).
FT SEQUENCE 2459 AA; 269497 MW; 2E3F6872DEDB8BA2 CRC64;
Query Match 11.7%; Score 81.5; DB 1; Length 2459;
Best Local Similarity 27.8%; Pred. No. 1.3e+02;
Matches 25; Conservative 18; Mismatches 28; Indels 19; Gaps 4;
Qy 48 LSKDTFKVPVK-KIEEKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSQK 106
Db 631 VTQKVKVKKEIKTPKEEKKEE--RPKKEVAKKED-----KTPLEKKEKPKK 674
Qy 107 SDSTKQDVATVTL--DKNNISSKSTTNPNK 134
Db 675 EAKKEIKKEIKKEEKKELKEVKEKTPLEK 704
```

```
RESULT 26
PAT1_YEAST
ID PAT1_YEAST STANDARD; PRT; 797 AA.
AC P25644;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Topoisomerase II-associated protein PAT1.
GN PAT1 OR YCR077C OR YCR77C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Ballesta J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,
  Sanz E.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS. AND CHARACTERIZATION.
RX MEDLINE=95373282; PubMed=7645349;
RA Rodriguez-Cousino N., Lill R., Neupert W., Court D.A.;
RT "Identification and initial characterization of the cytosolic protein
  Ycr77p."
RT Yeast 11:581-585 (1995).
```

```
RN [3]
RP FUNCTION.
RX MEDLINE=971128274; PubMed=8972867;
RA Wang X., Watt P.M., Louis E.J., Borts R.H., Hickson I.D.;
RT "Pat1, a topoisomerase II-associated protein required for faithful
  chromosome transmission in Saccharomyces cerevisiae."
RL Nucleic Acids Res. 24:4791-4797 (1996).
CC -1- FUNCTION: NECESSARY FOR ACCURATE CHROMOSOME TRANSMISSION DURING
  CELL DIVISION. INTERACTS WITH TOPOISOMERASE TOP2.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: TO S. POMBE SPBC19G7.10C.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announcement/
  or send an email to license@isb-sib.ch).
```

```
CC EMBL; X59720; CAA42266.1; -.
CC PIR; S53590; S53590.
CC SGD; S0000673; PAT1.
CC GO; GO:0005843; C:cytosolic small ribosomal subunit (sensu Eu. . .; IDA.
CC GO; GO:0007059; P:chromosome segregation; IMP.
CC GO; GO:0006446; P:regulation of translational initiation; IGI.
CC DOMAIN 195 200 POLY-PRO.
CC SEQUENCE 797 AA; 88487 MW; 2A54F39AE3E75ECE CRC64;
Query Match 11.6%; Score 81; DB 1; Length 797;
Best Local Similarity 25.8%; Pred. No. 45;
Matches 40; Conservative 21; Mismatches 58; Indels 36; Gaps 7;
Qy 5 STTVSE---EDRTLPVVK-----GELEKG-----YQFDGWEISGPGKKDAGTVI 46
Db 293 SQIVTEPDYNEDFYQVYKIQRGITSENKGLIARAYLEHSHGRGVRKRTD----- 347
Qy 47 NLSKDTFKVPVKKIEEKEENK-----PTFDVSKKNDPQVNHSQLNESHKRE 96
Db 348 -IALQRMQSQVERKAVTAKERPSKLDQQAAGNSQDNKQANTVLGKISLTLSKNRPR 406
Qy 97 DLQ--REHSQKSDSTKQDVATVLDKNNISSKSTT 129
Db 407 QLQIPROQSPSPDALKDVTDSLNTVDLSSGSSS 441
```

```
RESULT 27
YCG1_YEAST
ID YCG1_YEAST STANDARD; PRT; 853 AA.
AC P25588; P25589; P27513; P87003;
DT 01-MAY-1992 (Rel. 22, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 97.9 kDa protein in CHAI-KKRI intergenic region.
GN YCL061C OR YCL61C/YCL60C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Rasmussen S.W., von Wettstein D.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Gromadka R.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 384-596 FROM N.A.
RX STRAIN=ATCC 28383 / FL100;
RX MEDLINE=90384830; PubMed=2169608;
RA Kern L.;
RT "The URK1 gene of Saccharomyces cerevisiae encoding uridine kinase.";
```

RC STRAIN=972;  
RC MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Scourto J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidaoglou J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Welljans I., Vansteenkels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,

CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO  
CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO





RT "An essential yeast gene with homology to the exonuclease-encoding  
RT XRN1/XEM1 gene also encodes a protein with exoribonuclease  
RT activity.";  
RL Mol. Cell. Biol. 13:341-350(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93268292; PubMed=8497260;  
RA Aldrich T.L., di Segni G., McConaughy B.L., Keen N.J., Whelen S.,  
RA Hall B.D.;  
RT "Structure of the yeast TAP1 protein: dependence of transcription  
RT activation on the DNA context of the target gene.";  
RL Mol. Cell. Biol. 13:3434-3444(1993).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Landt O., Hiesel R., Unseid M.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Bohn C., Bolotin-Fukuhara M., Daignan-Fornier B., Dang D.V.,  
RA Valens M.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: May function in the processing and/or trafficking of  
CC nuclear mRNA. May be involved in general transcription as well.  
CC Possesses 5'-3' exoribonuclease activity degrading poly(A) to  
CC mainly 5'-AMP.  
CC -1- COFACTOR: Requires magnesium.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: Belongs to the 5'-3' exonuclease family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; S61567; AAB26818.1; -;  
CC EMBL; M95626; AAB34960.1; -;  
CC EMBL; L06011; AAA16950.1; -;  
CC EMBL; Z1746; -; NOT ANNOTATED\_CDS.  
CC EMBL; Z74956; CA99240.1; -;  
CC PIR; S20126; S20126.  
CC SGD; S0005574; RAT1.  
CC GO; GO:0005634; C:nucleus; IDA.  
CC GO; GO:0004534; F:5'-3' exoribonuclease activity; IDA.  
CC GO; GO:0006365; P:35S primary transcript processing; IMP.  
CC GO; GO:0006396; P:RNA processing; IMP.  
CC InterPro; IPR004859; Put\_53exo.  
CC Pfam; PF03159; XRN\_N; 1.  
CC Nuclear protein; Hydroxylase; Nuclease; Exonuclease; Repeat.  
FT DOMAIN 525 528 POLY-GLU.  
FT DOMAIN 793 797 POLY-ASN.  
FT DOMAIN 955 999 CONTAINS 2 X SRD, 2 X NNNY, AND 2 X  
FT YSGN REPEATS.  
FT MUTAGEN 683 Y->H: IN ALLELE TAP1-1; ACTIVATES  
FT TRANSCRIPTION OF THE PROMOTER-DEFECTIVE  
FT YEAST SUP4 TRNA(TYR) ALLELE SUP4A53T61.  
SQ SEQUENCE 1006 AA; 115933 MW; 5DD5B0245F3E12A CRC64;  
  
Query Match 11.6%; Score 80.5; DB 1; Length 1006;  
Best Local Similarity 20.0%; Pred. No. 62;  
Matches 34; Conservative 37; Mismatches 44; Indels 55; Gaps 9;  
  
Qy 1 KEMES-----TIVSEEDFILPVY--KGELEKGYQFGDGEISGFGKKGAGYVNLKSDT 52  
Db 413 KWSKSGDRHTVATEQ---LQWYDTQGNLAKG-----SWNLTTSD-----WVRLKKEL 458  
  
Qy 53 FI-----KPVFKKIEKKEENKPTTFDVSKKKDPQVNHQ-----LNESH 93  
Db 459 MLANEGNEEALAKVQKQSDKNELMKDISKEIIDAVSKANKTNFNLAEVVMKQKIINKKH 518  
Qy 94 RKDLQREEH-----SQKSUSTKDVATVLDK-----NNISSKSTT 129

Db 519 RLEKONDEEIEIAKSKVKTEKAESECDLDAETKDEIVADVNDRENSSETT 568  
  
RESULT 33  
LIP\_STAEP  
ID -LIP\_STAEP STANDARD; PRT; 688 AA.  
AC Q02510;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Lipase precursor (EC 3.1.1.3) (Glycerol ester hydrolase).  
GN GEHC OR S0281.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 303-315.  
RC STRAIN=9;  
RX MEDLINE=93171870; PubMed=8436947;  
RA Farrell A.M., Foster T.J., Holland K.T.;  
RT "Molecular analysis and expression of the lipase of Staphylococcus  
RT epidermidis.";  
RL J. Gen. Microbiol. 139:267-277(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
RA Chen Z., Wen Y.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a  
CC fatty acid anion.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS  
CC NEGATIVELY REGULATED BY BACTERIOPHAGE LYSOGENIZATION (LIPASE  
CC CONVERSION).  
CC -1- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M95577; AAA19729.1; -;  
CC EMBL; AE016744; AAO03878.1; -;  
CC PIR; A47705; A47705.  
CC InterPro; IPR005877; Gpos\_Y5IRK.  
CC InterPro; IPR000734; Lipase.  
CC InterPro; IPR000379; Ser\_estrs\_site.  
CC Pfam; PF04650; Y5IRK\_signal; 1.  
CC TIGRFAMs; TIGR01168; Y5IRK\_signal; 1.  
CC PROSITE; PS00120; LIPASE\_SER; 1.  
CC Hydroxylase; Lipid degradation; Zymogen; Signal; Complete proteome.  
FT SIGNAL 1 35 POTENTIAL.  
FT PROPEP 36 302 REMOVED IN THE MATURE FORM.  
FT CHAIN 303 688 LIPASE.  
FT ACT\_SITE 418 418 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 648 648 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT CONFLICT 96 96 W->L (IN REF. 1).  
FT CONFLICT 120 120 E->G (IN REF. 1).  
SQ SEQUENCE 688 AA; 77343 MW; 6C95DB3A78AF86F6 CRC64;  
  
Query Match 11.5%; Score 80; DB 1; Length 688;  
Best Local Similarity 26.6%; Pred. No. 46;  
Matches 34; Conservative 24; Mismatches 50; Indels 20; Gaps 6;  
  
Qy 11 EDFILPVYKGELEKGYQFGDGEISGFGKKGAGYVNLKSDTFFKVFVKIE-EKKEE 68  
Db 94 EDWI-----QQKQSDNKSNESVVEQNKENEAFVQNHSEE---KQOEQVELEKHA 144

69 NKPTEDVSKDNPOV--NHSQLNESHKEDLOREHSHKQSDTKDVTATVLDKNNISSK 126  
 145 NNQTLHSAAGNEDVKTPSQDLNAAQDSQENLSKQDTQSSKTTDL-----R 197  
 127 STTNPNK 134  
 198 ATAQNSK 205

RESULT 34  
 ID IF2P HUMAN STANDARD; PRT; 1220 AA.  
 AC O60841; O95805; Q9UR81; Q9UMN7;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Translation initiation factor IF-2.  
 GN IF2 OR KIAA0741.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99087487; PubMed=9872452;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RA "Prediction of the coding sequences of unidentified human genes. XI.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 5:277-286 (1998).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cervical Carcinoma;  
 RX MEDLINE=99362399; PubMed=10432305;  
 RA Wilson S.A., Siero-Vazquez C., Edwards N.J., Iourin O., Byles E.D.,  
 RA Kotsopoulos E., Adamson C.S., Kingsman S.M., Kingsman A.J.,  
 RA Martin-Rendon E.;  
 RA "Cloning and characterization of hIF2, a human homologue of bacterial  
 RT translation initiation factor 2 and its interaction with HIV-1  
 RT matrix.";  
 RL Biochem. J. 342:97-103 (1999).  
 [3]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE=Testis;  
 RX MEDLINE=99218282; PubMed=10200264;  
 RA Lee J.H., Choi S.K., Roll-Mecak A., Burley S.K., Dever T.E.;  
 RA "Universal conservation in translation initiation revealed by human  
 RT and archaeal homologs of bacterial translation initiation factor  
 RT IF2.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:4342-4347 (1999).  
 [4]  
 RP SEQUENCE OF 89-1220 FROM N.A.  
 RC TISSUE=Testis;  
 RA Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;  
 RA Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RP SEQUENCE OF 833-1220 FROM N.A.  
 RX MEDLINE=21064499; PubMed=11124703;  
 RA Stanchi F., Bertocco E., Toppo S., Dioguardi R., Simionati B.,  
 RA Cannata N., Zibello R., Lanfranchi G., Valle G.;  
 RA "Characterization of 16 novel human genes showing high similarity to  
 RT yeast sequences.";  
 RL Yeast 18:69-80 (2001).  
 CC -!- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING  
 CC THE BINDING OF THE FORMYL METHIONINE-TRNA TO RIBOSOMES. SEEMS TO  
 CC FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB018284; BAA34461.1; -  
 CC EMBL; AJ006776; CAB44357.1; -  
 CC EMBL; AF078035; AAD16006.1; -  
 CC EMBL; AL133563; CAB63717.1; -  
 CC EMBL; AJ006412; CAA07018.1; -  
 CC PIR; T43483; T43483.  
 DR MIN; G06086; -  
 DR GO; G0:0003743; F:translation initiation factor activity; NAS.  
 DR GO; G0:0006446; P:regulation of translational initiation; NAS.  
 DR InterPro; IPR000795; EF\_GTPbind.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR InterPro; IPR000178; IF2.  
 DR InterPro; IPR005225; Small GTP.  
 DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
 DR PRINTS; PR00315; ELONGATNFT.  
 DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
 DR PROSITE; PS01176; IF2; FALSE\_NEG.  
 KW Initiation factor; Protein biosynthesis; GTP-binding.  
 FT NP\_BIND 638 645 GTP (BY SIMILARITY).  
 FT DOMAIN 39 50 POLY-LYS.  
 FT DOMAIN 94 99 POLY-LYS.  
 FT DOMAIN 138 142 POLY-ASP.  
 FT DOMAIN 313 322 POLY-LYS.  
 FT DOMAIN 353 356 POLY-GLU.  
 FT DOMAIN 361 364 POLY-GLU.  
 FT DOMAIN 491 496 POLY-GLU.  
 FT DOMAIN 529 567 ASP/GLU-RICH (ACIDIC).  
 FT MUTAGEN 640 640 V->G: LOSS OF ACTIVITY IN VIVO. RETAINS  
 FT MUTAGEN 706 706 FULL ACTIVITY IN VITRO.  
 FT MUTAGEN 706 706 H->E: LOSS OF ACTIVITY; BOTH IN VIVO AND  
 FT MUTAGEN 759 759 H->Q: LOSS OF ACTIVITY IN VIVO. PARTIAL  
 FT MUTAGEN 759 759 D->N: LOSS OF ACTIVITY; BOTH IN VIVO AND  
 FT CONFLICT 64 64 IN VITRO.  
 FT CONFLICT 92 92 E -> G (IN REF. 1).  
 FT CONFLICT 180 180 T -> I (IN REF. 2).  
 FT CONFLICT 256 256 I -> M (IN REF. 3).  
 FT CONFLICT 522 522 K -> R (IN REF. 3).  
 FT CONFLICT 549 549 T -> K (IN REF. 1).  
 FT CONFLICT 549 549 E -> V (IN REF. 3).  
 FT CONFLICT 669 669 G -> W (IN REF. 3).  
 FT CONFLICT 894 894 E -> K (IN REF. 2).  
 SQ SEQUENCE 1220 AA; 138799 MW; 010CC15DE475B5B8 CRC64;  
 Query Match 11.5%; Score 80; DB 1; Length 1220;  
 Best Local Similarity 24.8%; Pred. No. 83;  
 Matches 35; Conservative 19; Mismatches 63; Indels 24; Gaps 4;  
 QY 10 BEDFILPVYKGELEKGVQFDGWEISGEGKKDAGVINLSKDTFKPKVPFKLEEKKESEN 69  
 DB 189 ESDEFLOSRRGQKQKNQKNKPGFNIESGNEDDASFK-----IKTVAQKKAERER 240  
 QY 70 KPTFD---VSKKXNDPNVHNSQLNESHKEDLOREHSHKQSDTKDVTATVLDKNNISSK 113  
 DB 241 KKRDEEKAKRLKKEKELETKGKQDSQKESQKRFEEETVKSQKTVDTGTVPASEEKAE 300  
 QY 114 TATVLDKNNISSKSTTNPNK 134  
 DB 301 TPTAAEDDNEGDKKKDKKKK 321  
 RESULT 35  
 ID CENC HUMAN STANDARD; PRT; 943 AA.  
 AC Q03188; Q9P0M5;  
 DT 01-OCT-1993 (Rel. 27, Created)



01-OCT-1993 (Rel. 27, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 Centromere protein C (CENP-C) (Centromere autoantigen C).  
 CENPC1 OR CENPC.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=92323541; PubMed=1339310;  
 RA Satoh H., Tomkiel J., Cooke C.A., Ratnie H. III, Maurer M.,  
 Rothfield N.F., Earnshaw W.C.;  
 RT "CENP-C, an autoantigen in scleroderma, is a component of the human  
 inner kinetochore plate.";  
 RL Cell 70:115-125(1992).  
 RN [2]  
 RP SEQUENCE OF 1-21 FROM N.A.  
 RA Poppe M., Botz J., Hamm B., Dobat K., Eickelbaum W., Pawletz N.,  
 Arand M., Knehr M.;  
 RT "Promoter characterization of centromere protein C reveals its  
 participation in cell cycle regulation in late G1-phase and expression  
 control by E2F-1, p107 and Sp-1.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Component of the inner kinetochore plate. Required for  
 normal kinetochore assembly.  
 CC -!- SUBUNIT: Binds to DAXX.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; M95724; AAA51974.1; -;  
 DR EMBL; AF151723; AAF73191.1; -;  
 DR PIR; A42681; A42681.  
 DR Genew; HGNC:1854; CENPC1.  
 DR MIM; 117141; -;  
 DR GO; GO:0005699; C:kinetochore; TAS.  
 DR GO; GO:0003677; F:DNA binding activity; TAS.  
 DR InterPro; IPR007113; Cupin\_sup.  
 DR Chromosomal protein; Nuclear protein; DNA-binding; Phosphorylation;  
 KW Centromere.  
 FT DOMAIN 259 273 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 484 499 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 558 574 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 780 798 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD RES 75 75 PHOSPHORYLATION (POTENTIAL).  
 FT MOD RES 732 732 PHOSPHORYLATION (POTENTIAL).  
 SQ SEQUENCE 943 AA; 106925 MW; 6D80810A3E476376 CRC64;  
 Query Match 11.4%; Score 79.5; DB 1; Length 943;  
 Best Local Similarity 23.1%; Pred. No. 69;  
 Matches 36; Conservative 16; Mismatches 55; Indels 49; Gaps 5;  
 QY 16 PVYGELEKRGQFDGWEISGFEKKDAGVNLNSKDTIKPVFKIEKK-----ESEN 69  
 DB 375 PSDKTLVLTSLALIDETVNNYRSTKYEMYSKNAEKSFKRTIKQKQKRWKPAEQQ- 433  
 QY 70 KPTFDVSKKDD-NPQVNHSQLNSHRKEDLQREH-----SOKSDSTK 111  
 DB 434 ---LDVQSGKDNIEHTSHITQDEFORNSDRNMEHEEENGVNDCVSKQPPVGVSKSKSTRK 490  
 QY 112 D-----VYATVLDKNNISK 126  
 DB 491 DKESKKKRFSESKNKLVPBEVTSVTVKRSISR 526

RESULT 36  
 TRI DROME STANDARD; PRT; 268 AA.  
 AC P36188; Q9VMVY1; Q9VMVY3; Q9VMVY4;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE Troponin I (TNI) (Wings apart-A protein) (Heldup protein).  
 GN WUPA OR HDP OR TNI OR CG7178.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OX Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 9), AND FUNCTION.  
 RC STRAIN=Canton-S; TISSUE=Embryo, and Larva;  
 RX MEDLINE=91115093; PubMed=1899228;  
 RA Barbas J.A., Gaiceran J., Krah-Jentgens I., de la Pompa J.L.,  
 Canal I., Pongs O., Ferrus A.;  
 RT "Troponin I is encoded in the haplolethal region of the Shaker gene  
 complex of Drosophila.";  
 RL Genes Dev. 5:132-140(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 6).  
 RC STRAIN=Oregon-R; TISSUE=Pupae;  
 RX MEDLINE=91340840; PubMed=1908472;  
 RA Beall C.J., Fyrberg E.;  
 RT "Muscle abnormalities in Drosophila melanogaster heldup mutants are  
 caused by missing or aberrant troponin-I isoforms.";  
 RL J. Cell Biol. 114:941-951(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,  
 Foaler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu J.C.;  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).





```
AC Q23917;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3',5'-cyclic-nucleotide phosphodiesterase regA (EC 3.1.4.17) (PDEase
DE regA).
GN REGA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Shaelsky G., Escalante R., Loomis W.F.;
RX "Developmental signal transduction pathways uncovered by genetic
RT suppressors.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX2;
RA Thomason P.A., Traynor D., Cavet G., Chang W.T., Harwood A.J.,
RA Kay R.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MORPHOLOGICAL SUPPRESSOR OF TAGS.
CC -!- CATALYTIC ACTIVITY: Nucleoside 3',5'-cyclic phosphate + H(2)O =
CC nucleoside 5'-phosphate.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT LOW LEVELS IN VEGETATIVE CELLS
CC AND AT HIGH LEVELS IN PRESPORE AND PRETALK CELLS DURING
CC DEVELOPMENT.
CC -!- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC FAMILY.
CC -!- SIMILARITY: Contains 1 response regulatory domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U60170; AAB03508.1; -.
CC EMBL; AJ005398; CAA06513.1; -.
CC Dictyob; DD02055; regA.
CC InterPro; IPR003607; Met_phosphohydro.
CC InterPro; IPR002073; PDEase.
CC Pfam; PF00233; PDEase; 1.
CC ProDom; PD000039; Response_reg.
CC PRINTS; PR00387; PDIESTERASE1.
CC ProDom; PD000039; Response_reg; 1.
CC SMART; SM00471; Hdc; 1.
CC SMART; SM00448; REC; 1.
CC PROSITE; PS01126; PDEASE_I; 1.
CC PROSITE; PS01110; RESPONSE_REGULATORY; 1.
CC Hydrolase; CAMP; CGNP.
FT DOMAIN 161 280 RESPONSE REGULATORY.
FT DOMAIN 18 28 POLY-SER.
FT DOMAIN 52 69 POLY-ASN.
FT DOMAIN 87 96 POLY-THR.
FT DOMAIN 100 121 POLY-ASN.
FT DOMAIN 166 169 POLY-ASP.
FT DOMAIN 770 776 POLY-SER.
SQ SEQUENCE 793 AA; 91175 MW; 65065A96620F8C27 CRC64;
Query Match 11.4%; Score 79; DB 1; Length 793;
Best Local Similarity 24.4%; Pred. No. 63;
Matches 21; Conservative 19; Mismatches 34; Indels 12; Gaps 2;
QY 52 TFIRPVFKKIB-----EKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSQKS 107
DB 33 TSLKSMISGIEINLVHSGKNDN-----KNNNNNNNNNNNNKQKQDIVSLNNSSSS 84
```

```
QY 108 DSTKDVATVTLDKNNISSKSTNNPN 133
DB 85 NNITTTTTTTTNSNNNNNNNNNN 110
RESULT 38
Y372 AQUAE STANDARD; PRT; 531 AA.
AC O6695;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_372.
GN AQ_372.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., AuJay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE00686; AAC06652.1; -.
CC PIR; C70333; C70333.
CC Hypothetical protein; Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 531 AA; 62830 MW; 6499DFE803C2F7CF CRC64;
Query Match 11.3%; Score 78.5; DB 1; Length 531;
Best Local Similarity 25.3%; Pred. No. 45;
Matches 25; Conservative 17; Mismatches 34; Indels 23; Gaps 4;
QY 3 MSSTIVSEDFILPVYKGELEKGYQFDGWEISGEGKKDAGYVNLKDTFIKPVFKKIE 62
DB 66 MKEQILSKKLSIETLKNRLNPNYTF-----INGYD-----IISKFYRNSIF---- 107
QY 63 EKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQRE 101
DB 108 ----ANNPYKDI--KOENSKLLNLIINDIHRKEDMDYE 141
RESULT 39
CYLL HUMAN STANDARD; PRT; 598 AA.
AC P35663;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cyclicin I (Multiple-band polypeptide I) (Fragment).
GN CYLCI OR CYLI OR CYL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=93359502; PubMed=8354692;
RA Hess H., Heid H., Franke W.W.;
```

```
RT "Molecular characterization of mammalian cylicin, a basic protein of
RT the sperm head cytoskeleton."
RL J. Cell Biol. 122:1043-1052(1993).
CC -!- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
CC BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC -!- TISSUE SPECIFICITY: Testis.
CC -!- DEVELOPMENTAL STAGE: SPECIFIC TO LATE SPERMATOGENESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z22780; CAA80457.1; -
DR PIR; B40713; B40713.
DR Genew; HGNC:2582; CYLC1.
DR MIM; 603121; -
DR GO; GO:0005856; C:cytoskeleton; NAS.
DR GO; GO:0005198; F:structural molecule activity; NAS.
DR GO; GO:0007283; P:spermatogenesis; NAS.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT NON_TER 1 1
FT DOMAIN 479 500 8 APPROXIMATE TANDEM REPEATS.
FT REPEAT 225 252 1.
FT REPEAT 253 289 2.
FT REPEAT 290 326 3.
FT REPEAT 327 364 4.
FT REPEAT 365 400 5.
FT REPEAT 401 438 6.
FT REPEAT 439 478 7.
FT REPEAT 479 500 8.
FT DOMAIN 548 598 PRO-RICH.
FT SEQUENCE 598 AA; 68034 MW; 062BA2E2D2AB61F7 CRC64;
SQ
Query Match 11.3%; Score 78.5; DB 1; Length 598;
Best Local Similarity 22.7%; Pred. No. 51;
Matches 32; Conservative 28; Mismatches 60; Indels 21; Gaps 8;
QY 5 STIVSEE-----DFILPVYKGELEKGYQDQWEISGFEKGKQAGYVINLS-----KDTFIKP 56
DB 182 SEICSENSLVDFLMVGQSD-DESGINFDW-LRNYSQNNKSNYSKTKYTKTKDT--KK 237
QY 57 VFKKIEEKKEENKPTFDVSKKNDPNVNHQSOLNESHKEDLQREHS-----QKSDSTKDV 113
DB 238 NAKKSSDAEESDK-----DAKDKKKYK-KNVKDDKKDKVKKDTESDAGSDKDER 291
QY 114 TATVLDKNNISSKSTNNPNK 134
DB 292 KDTKDKKKLKKDKKKDKTK 312
RESULT 40
DNL1 CABEL STANDARD; PRT; 847 AA.
AC Q27474;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA ligase (EC 6.5.1.1) (Polydeoxyribonucleotide synthase [ATP]).
GN LIG-1 OR C29A12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilkinson J.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
```

```
CC -!- FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA
CC RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA.
CC -!- CATALYTIC ACTIVITY: ATP + {deoxyribonucleotide}(N) +
CC {deoxyribonucleotide}(M) = AMP + diphosphate +
CC {deoxyribonucleotide}(N+M).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z73970; CAA98242.1; -
DR PIR; T19544; T19544.
DR WormPep; C29A12.3; CE05328.
DR InterPro; IPR000977; DNA_ligase.
DR Pfam; PF01068; DNA_ligase_1.
DR Pfam; PF04679; DNA_ligase_A_C; 1.
DR Pfam; PF04675; DNA_ligase_A_N; 1.
DR TIGRFAMs; TIGR00574; dnll; 1.
DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
DR PROSITE; PS00333; DNA_LIGASE_A2; 1.
DR PROSITE; PS00160; DNA_LIGASE_A3; 1.
KW DNA repair; DNA replication; DNA recombination; Cell division; Ligase;
KW ATP-binding.
FT BINDING 363 363 AMP (BY SIMILARITY).
SQ SEQUENCE 847 AA; 94501 MW; A32B7D279F91002F CRC64;
Query Match 11.3%; Score 78.5; DB 1; Length 847;
Best Local Similarity 25.2%; Pred. No. 74;
Matches 37; Conservative 24; Mismatches 59; Indels 27; Gaps 6;
QY 2 EMSSTIVSEEDFILPVYKGE-----LEKGYQDQWEISGFEKGKQAGYVINL----- 48
DB 674 DKNSDDATSSQVLEMYKQAEAFANQIEKADVD--EDDFEKEDEEEELNMTNVSEG 731
QY 49 -SKDTFTKVPKPKIEEKKEENKPTFDVSKKK--DNQVNHQSOLNESHKEDLQREHSQ 105
DB 732 SSKENPVK-----EIKKETPK--SVSPKPKFEKPPVKSSPVNKPSPKSPKKEAEK 782
QY 106 KSDSTKDV TATVLDKNNISSKSTNNP 132
DB 783 KKGVPASIFSSSTTKKNEKDKVESPSP 809
Search completed: February 10, 2004, 10:49:50
Job time : 7.22948 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 23.725 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_640\_773

Perfect score: 696

Sequence: 1 KEMSTTVSEEDFILPVYK.....ATVLKKNISSKSTNNPNK 134

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:\*

- 1: sp\_archea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	696	100.0	2119	2 Q9AHT5	Q9Aht5 streptococc
2	696	100.0	2140	16 Q97RY6	Q97ry6 streptococc
3	693	99.6	2144	2 Q9S4M8	Q9s4m8 streptococc
4	693	99.6	2144	16 Q8DOP7	Q8dop7 streptococc
5	112	16.1	361	5 Q9SP15	Q9sp15 plasmodium
6	110	15.8	346	5 Q9U0G0	Q9u0g0 plasmodium
7	110	15.8	379	5 Q25705	Q25705 plasmodium
8	109	15.7	3008	5 Q81436	Q81436 plasmodium
9	108.5	15.6	600	5 Q77355	Q77355 plasmodium
10	108.5	15.6	1038	13 Q90784	Q90784 gallus gall
11	107	15.4	354	5 Q25995	Q25995 plasmodium
12	107	15.4	354	5 Q81J55	Q81j55 plasmodium
13	107	15.4	829	5 Q815F3	Q815f3 plasmodium
14	103.5	14.9	379	5 Q9U6C4	Q9u6c4 plasmodium
15	103.5	14.9	380	5 Q26019	Q26019 plasmodium
16	102.5	14.7	379	5 Q25706	Q25706 plasmodium

17	101.5	14.6	382	5	Q9V7J0	Q9v7j0 drosophila
18	101.5	14.6	556	5	Q9V7I9	Q9v7i9 drosophila
19	101.5	14.6	785	5	Q9GQ82	Q9gq82 drosophila
20	100	14.4	329	5	Q9NFV9	Q9nf9v plasmodium
21	100	14.4	1130	5	Q8IJ24	Q8ij24 plasmodium
22	99	14.2	211	5	P91488	P91488 caenorhabdi
23	99	14.2	1859	5	Q8IC27	Q8ic27 plasmodium
24	99	14.2	2563	5	Q8I3A0	Q8i3a0 plasmodium
25	98	14.1	157	5	Q9VQV0	Q9vqv0 drosophila
26	98	14.1	951	5	Q96229	Q96229 plasmodium
27	97.5	14.0	449	5	Q8IHW3	Q8ihw3 plasmodium
28	97.5	14.0	556	5	Q9S893	Q9s893 drosophila
29	95	13.6	622	16	Q98QA1	Q98qa1 mycoplasma
30	95	13.6	3026	5	Q8ILS9	Q8ils9 plasmodium
31	95	13.6	3193	5	Q8I590	Q8i590 plasmodium
32	94.5	13.6	312	16	Q9PPL5	Q9ppl5 campylobact
33	94.5	13.6	614	5	Q8TIX2	Q8tix2 dictyosteli
34	94	13.5	210	5	Q02124	Q02124 caenorhabdi
35	94	13.5	535	5	Q17595	Q17595 caenorhabdi
36	94	13.5	1434	5	Q8IJT3	Q8ijj3 plasmodium
37	94	13.5	3455	5	Q8IKQ1	Q8ikq1 plasmodium
38	93	13.4	450	2	Q52006	Q52006 streptococc
39	93	13.4	455	5	Q8MN05	Q8mn05 dictyosteli
40	93	13.4	3127	5	Q8IDA0	Q8ida0 plasmodium
41	92.5	13.3	219	16	Q9XOM6	Q9xom6 thermotoga
42	92.5	13.3	325	5	O44016	O44016 dictyosteli
43	92.5	13.3	421	16	Q8EJQ7	Q8ejq7 shewanella
44	92.5	13.3	426	5	Q96300	Q96300 dictyosteli
45	92.5	13.3	1183	5	Q8IC11	Q8ic11 plasmodium

#### ALIGNMENTS

RESULT 1

Q9AHT5	ID	Q9AHT5	PRELIMINARY;	PRT;	2119 AA.
AC	Q9AHT5;				
DT	01-JUN-2001	(TRENBLrel. 17, Created)			
DT	01-JUN-2001	(TRENBLrel. 17, Last sequence update)			
DT	01-MAR-2003	(TRENBLrel. 23, Last annotation update)			
DE	Serine protease (Fragment).				
GN	PTA.				
OS	Streptococcus pneumoniae.				
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
OC	Streptococcus.				
OX	NCBI_TaxID=1313;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=N4;				
RX	MEDLINE=2116976; PubMed=11179332;				
RA	Wizemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,				
RA	Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,				
RA	Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,				
RA	Langermann S., Johnson S., Koenig S.;				
RT	"Use of a Whole Genome Approach To Identify Vaccine Molecules				
RT	Affording Protection Against Streptococcus pneumoniae Infection."				
RL	Infect. Immun. 69:1593-1598(2001).				
CC	-1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY				
CC	AN AMIDE BOND (BY SIMILARITY).				
DR	EMBL; AF291699; AAK19159.1; -				
DR	HSSP; P00782; 2SPT.				
DR	MEROPS; S08.064; -				
DR	InterPro; IPR001899; Gram_pos_anchor.				
DR	InterPro; IPR006192; LPXTG.				
DR	InterPro; IPR003137; PA.				
DR	InterPro; IPR00209; Peptidase_S8.				
DR	InterPro; IPR001680; WD40.				
DR	Pfam; PF00746; Gram_pos_anchor; 1.				
DR	Pfam; PF02225; PA; 1.				
DR	Pfam; PF00082; Peptidase_S8; 2.				
DR	PRINTS; PR00723; SUBTILISIN.				
DR	TIGRFAMs; TIGR01167; LPXTG_anchor; 1.				

DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Cell wall; Peptidoglycan-anchor; Protease.  
FT NON\_TER 1  
SQ SEQUENCE 2119 AA; 238226 MW; 51799B7F6B960A6A CRC64;  
  
Query Match 100.0%; Score 696; DB 2; Length 2119;  
Best Local Similarity 100.0%; Pred. No. 1.4e-45;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVILNSKDTFIKPVFKK 60  
DB 1952 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVILNSKDTFIKPVFKK 2011  
  
QY 61 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120  
DB 2012 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2071  
  
QY 121 NNISKSSTNNPNK 134  
DB 2072 NNISKSSTNNPNK 2085  
  
RESULT 2  
Q97RY6 PRELIMINARY; PRT; 2140 AA.  
AC Q97RY6;  
DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Serine protease, subtilase family.  
GN SP0641.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1] SEQUENCE FROM N.A.  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21357209; PubMed=11463916;  
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Unavay L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
RA Holtzaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae";  
RL Science 293:498-506(2001).  
DR EMBL; AE007373; AA74791.1; -.  
DR TIGR; SP0641; -.  
DR MEROPS; S08.064; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR PRINTS; PR00723; SUBTILISIN.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Protease; Complete proteome.

SQ SEQUENCE 2140 AA; 240426 MW; FA44AD9E2938B334 CRC64;  
  
Query Match 100.0%; Score 696; DB 16; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 1.4e-45;  
Matches 134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVILNSKDTFIKPVFKK 60  
DB 1973 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVILNSKDTFIKPVFKK 2032  
  
QY 61 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 120  
DB 2033 IEKKEEENKPTFDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDK 2092  
  
QY 121 NNISKSSTNNPNK 134  
DB 2093 NNISKSSTNNPNK 2106  
  
RESULT 3  
Q9S4M8 PRELIMINARY; PRT; 2144 AA.  
AC Q9S4M8;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Cell wall-associated serine proteinase PrtA precursor.  
GN PrtA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1] SEQUENCE FROM N.A.  
RP SEQUENCE FROM N.A.  
RX STRAIN=3.B;  
RA Bethe G., ten Thoren E., Bongaerts R.J.M., Heinz H.-P., Zysk G.;  
RT "Cloning and sequencing of a novel surface protease of Streptococcus pneumoniae";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY AN AMIDE BOND (BY SIMILARITY).  
DR EMBL; AF127143; AAD48399.1; -.  
DR HSP; P00782; 2SPT.  
DR MEROPS; S08.064; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Cell wall; Peptidoglycan-anchor; Signal.  
FT SIGNAL 19  
FT CHAIN 20 2144  
FT CELL\_WALL-ASSOCIATED SERINE PROTEINASE  
FT PrtA.  
SQ SEQUENCE 2144 AA; 240724 MW; 20525511470741331 CRC64;  
  
Query Match 99.6%; Score 693; DB 2; Length 2144;  
Best Local Similarity 99.3%; Pred. No. 2.4e-45;  
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVILNSKDTFIKPVFKK 60  
DB 1977 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGKDGAGYVILNSKDTFIKPVFKK 2036

```

Qy 61 IEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 120
Db 2037 IEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 2096

Qy 121 NNISKSSTNNPNK 134
Db 2097 NNISKSSTNNPNK 2110

RESULT 4
QSDQP7
ID QSDQP7 PRELIMINARY; PRT; 2144 AA.
AC QSDQP7;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN PrtA OR SP0361.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008434; AAK9365.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

Query Match 99.6%; Score 693; DB 16; Length 2144;
Best Local Similarity 99.3%; Pred. No. 2.4e-45;
Matches 133; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFIKPVFKK 60
Db 1977 KEMSTIVSEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVNLKDTFIKPVFKK 2036

Qy 61 IEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 120
Db 2037 IEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATVLDK 2096

Qy 121 NNISKSSTNNPNK 134
Db 2097 NNISKSSTNNPNK 2110

RESULT 5
Q95PI5
ID Q95PI5 PRELIMINARY; PRT; 361 AA.
AC Q95PI5;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Merozoite surface protein 3 (fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVO;
RA Hiseeda H., Saul A., Long C.A., Miller L.H., Stowers A.W.;
RT "Merozoite Surface Protein 3 and Protection Against Malaria in Aotus

```

```

RT Monkeys.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY044180; AAK94780.1; -.
FT NON_TER 1
FT NON_TER 361
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

Query Match 16.1%; Score 112; DB 5; Length 361;
Best Local Similarity 23.9%; Pred. No. 0.41;
Matches 37; Conservative 28; Mismatches 48; Indels 42; Gaps 6;

Qy 18 YKGELEKGYQ-----FD---GWEISGF--EGKDGAG-----YVI 46
Db 165 YAGKVEKDYERAKNAYQKANQAVLKAKEASSYDILGWFEFGGVPHEKKEENMLSHLYVS 224
Qy 47 NLSKDTFIKPVFKKIEKKEENKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOK 106
Db 225 SKDKENISKENDVDLDE-KEEAEETEELKEKNEETETSEISEDEEEEEEEEEEE 283
Qy 107 SDSTKDVATVLDKN-----NISSKSTNN 131
Db 284 NDKKKEQEKESNNENNDQKKDMEANLISKNNNN 318

RESULT 6
Q9U0G0
ID Q9U0G0 PRELIMINARY; PRT; 346 AA.
AC Q9U0G0;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Merozoite surface protein 3 (fragment).
GN MSP3.
OS Plasmodium reichenowi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5854;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20416497; PubMed=10960178;
RA Okenu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in
RT Plasmodium reichenowi and Plasmodium falciparum.";
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL; AJ252286; CAB65754.1; -.
FT NON_TER 1
FT NON_TER 346
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAPA010 CRC64;

Query Match 15.8%; Score 110; DB 5; Length 346;
Best Local Similarity 26.2%; Pred. No. 0.56;
Matches 34; Conservative 26; Mismatches 54; Indels 16; Gaps 5;

Qy 17 VYKGELEKGYQFD-GWEISGF--EGKDGAG-----YVINLSKDTFIKPVFKKIEKKEE 68
Db 184 VLKAKEASSYNYILGWFEFGGVPHEKKEENMLSHLYVSSKDKENISKENDVDLDE-KEE 242
Qy 69 NKPTFDVSKKNDPQVNHSQLNESHKEDLQREHSOKSDSTKDVATV-----LDKN 121
Db 243 AEETGEQELEKNEETESEINEDEQEKEEEEEKEEENNDKKEQAKESNDQKEDMEAQ 302

Qy 122 NISSKSTNN 131
Db 303 NLISKNNNN 312

RESULT 7
ID Q25705 PRELIMINARY; PRT; 379 AA.
AC Q25705;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Polymorphic antigen.

```

```
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=98156743; PubMed=9497029;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; U08851; AAC47831.1; -
SQ SEQUENCE 379 AA; 4334 MW; DCTAF106887C8AA0 CRC64;

Query Match 15.8%; Score 110; DB 5; Length 379;
Best Local Similarity 23.4%; Pred. No. 0.62;
Matches 37; Conservative 30; Mismatches 41; Indels 50; Gaps 7;

QY 18 YKLEKGYQ-----FD---GWEISGF--EGKKGAG-----YVI 46
Db 185 YAGKVEKDYERAKNAYQKANQAVLKAKAEASSYDYILGMEFGGVPEHKKEENMLSHLYVS 244
QY 47 NLSKDTFIPVFKTEEEKKEENKPTFDVSKKDNQVNHSQLN-----BSH 93
Db 245 SKDENISKENDVDLDE--KEEAETEEEEKKEETESISEDEEEKEEKEENE 303

QY 94 RKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNN 131
Db 304 KKKEQKEQSNENNQKDMEA-----QNLISKNNQNN 336

RESULT 8
Q8I436 PRELIMINARY; PRT; 3008 AA.
AC Q8I436;
DT 01-WAR-2003 (TrEMBLrel. 23, Created)
DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GE PFE0325W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RA Devlin K., Baker S., Davies P., Mungal K., Berrihan M., Pain A.,
RA Hall N., Bowman S., Churher C., Quail M., Barrell B.;
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berrihan M., Churher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C., Corton C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929351; CAD51431.1; -
KW Hypothetical protein.
SQ SEQUENCE 3008 AA; 356023 MW; 60BCBBEE15C599B4 CRC64;

Query Match 15.7%; Score 109; DB 5; Length 3008;
Best Local Similarity 23.4%; Pred. No. 0.62;
Matches 37; Conservative 30; Mismatches 41; Indels 50; Gaps 7;
```

```
Best Local Similarity 32.4%; Pred. No. 6.4;
Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

QY 44 YVINLSK----DTFIPVFKIEKEEENKPTFDVSKKDNQVNHSQLN-----LNSHHRKE 96
Db 2310 YDIELSKIEKFGASIGPVFTD--EENKKEENKN--EVNKEENKKEENKKEENKKEENKKE 2366
QY 97 DLQREH-----SQKSDSTKQVATVLDKNNISSK-----STTNNPNK 134
Db 2367 ENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENK 2414

RESULT 9
O77355 PRELIMINARY; PRT; 600 AA.
AC O77355;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 71.7 kDa protein.
GN PFC0465C, MAL3P4.20.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=3D7;
RC MEDLINE=99376085; PubMed=10488855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jaessal B., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RL falciparum";
RL Nature 400:532-538(1999).
DR EMBL; AL008970; CAA15610.2; -
DR InterPro; IPR002483; PWI.
DR Pfam; PF01480; PWI; 1.
DR SMART; SM00311; PWI; 1.
KW Hypothetical protein.
SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;

Query Match 15.6%; Score 108.5; DB 5; Length 600;
Best Local Similarity 29.3%; Pred. No. 1.3;
Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;

QY 7 IVSEDFILPVY----KGELEKGYQFDGWEISGFEGKK-----DAGYVINLSKDTFIPKV 57
Db 60 ILGFEDDILYECISQLKQSEKK--DGEEDKYLNAKKILNTGFIGNKKSDIFIEEL 116
QY 58 FKXI--EEKKEE-----ENKPTFDVSK-KKDNQVNHSQLN-----SHRK 95
Db 117 LELLINEEKEEHIADTLNENK-INDIKVKNENINENYNNENKDNKDKHEVSHQN 175
QY 96 E-----DLQREH-----SQKSDSTK-----DVTATVLDKNNISSKSTNN 130
Db 176 EHNINNNVLKKEKEYTDIQRDKRKHRSLSQKSDSKYKPKPKRKTSTIER-SLSNRYDE 234
QY 131 NPNK 134
Db 235 KTNK 238

RESULT 10
Q90784 PRELIMINARY; PRT; 1038 AA.
AC Q90784;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
```

```
DE Claustrian.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_taxid=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Cole G.J.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-451 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94157526; PubMed=7906711;
RA Burg M.A., Cole G.J.;
RT "Claustrian, an adhesive neural keratan sulfate proteoglycan, is
structurally related to MABP.";
RL J. Neurobiol. 25:1-22(1994).
DR EMBL; X67778; CAA47988.1; -.
SQ SEQUENCE 1038 AA; 117112 MW; 213D694A5B510927 CRC64;

Query Match 15.6%; Score 108.5; DB 13; Length 1038;
Best Local Similarity 28.6%; Pred. No. 2.3;
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

Qy 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKDGAGYVINLSKDTFTKPVFK- 59
Db 595 KPETKTIIVAEKDV-----TTKEQIGKSETSEKQASEKQDVKPKVTKSEKVKVEKA 646

Qy 60 KIEKKEBENKPTFDVSKKNDPQVNHSQLNESHKEDLQRE-----EHSQKSDSTKDV 113
Db 647 KPEKKDKERPKKESVSKKEKPLI---KKEKPKEDIKKEVKVEKKEKKEKKEV 702

RESULT 11
Q25995 ID Q25995 PRELIMINARY; PRT; 354 AA.
AC Q25995;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Antigen.
OS Plasmodium falciparum (isolate NF54).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_taxid=5843;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RX MEDLINE=95198774; PubMed=7891748;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT "Molecular variation in a novel polymorphic antigen associated with
Plasmodium falciparum merozoites.";
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NF54;
RX MEDLINE=98156743; PubMed=9497029;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
Plasmodium falciparum merozoite surface protein-3 (MSP-3)."
RL Mol. Biochem. Parasitol. 90:21-31(1997).
DR EMBL; L28825; AAC09377.1; -.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 15.4%; Score 107; DB 5; Length 354;
Best Local Similarity 23.8%; Pred. No. 0.98;
Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;

Qy 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 53
Db 185 KEASS-----YDYIL-----GWFGGVPFHKKEENMLSHLYVSSKDKENI 225

RESULT 12
Q81J55 ID Q81J55 PRELIMINARY; PRT; 354 AA.
AC Q81J55;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Merozoite surface protein 3.
GN PF10_0345.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_taxid=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014834; AAN35542.1; -.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 15.4%; Score 107; DB 5; Length 354;
Best Local Similarity 23.8%; Pred. No. 0.98;
Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;

Qy 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 53
Db 185 KEASS-----YDYIL-----GWFGGVPFHKKEENMLSHLYVSSKDKENI 225

Qy 54 IKPVFKKIEKKCE-----ENKPTFDVSKKNDPQVNHSQLNESHKEDLQ 100
Db 226 SKENDVDLDEKEEAEETEEELKEEKEETEESEISEDEEEEEEKEEENDKKQEK 285

RESULT 13
Q815F3 ID Q815F3 PRELIMINARY; PRT; 829 AA.
AC Q815F3;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFI1275C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_taxid=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
```

```
Qy 54 IKPVFKKIEKKCE-----ENKPTFDVSKKNDPQVNHSQLNESHKEDLQ 100
Db 226 SKENDVDLDEKEEAEETEEELKEEKEETEESEISEDEEEEEEKEEENDKKQEK 285

Qy 101 EHSQKSDSTKDVTTATVLDKNNISSKSTTN 131
Db 286 EQSNENDQKDMA-----QNLISKNNNN 311

RESULT 12
Q81J55 ID Q81J55 PRELIMINARY; PRT; 354 AA.
AC Q81J55;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Merozoite surface protein 3.
GN PF10_0345.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_taxid=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Perteira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McPadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014834; AAN35542.1; -.
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 15.4%; Score 107; DB 5; Length 354;
Best Local Similarity 23.8%; Pred. No. 0.98;
Matches 36; Conservative 28; Mismatches 43; Indels 44; Gaps 6;

Qy 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGF--EGKKDAG-----YVINLSKDTF 53
Db 185 KEASS-----YDYIL-----GWFGGVPFHKKEENMLSHLYVSSKDKENI 225

Qy 54 IKPVFKKIEKKCE-----ENKPTFDVSKKNDPQVNHSQLNESHKEDLQ 100
Db 226 SKENDVDLDEKEEAEETEEELKEEKEETEESEISEDEEEEEEKEEENDKKQEK 285

Qy 101 EHSQKSDSTKDVTTATVLDKNNISSKSTTN 131
Db 286 EQSNENDQKDMA-----QNLISKNNNN 311

RESULT 13
Q815F3 ID Q815F3 PRELIMINARY; PRT; 829 AA.
AC Q815F3;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFI1275C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_taxid=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
```







```

Query Match      14.7%; Score 102.5; DB 5; Length 379;
Best Local Similarity 25.2%; Pred. No. 2.4;
Matches 38; Conservative 27; Mismatches 41; Indels 45; Gaps 7;

QY 1 KEMSTVISEEDFILPVVKGLEKGYQDGEISGF--EGKKDAG-----YVINLSKDTF 53
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 KEASS-----YDIIL-----GMFEGGVPEHKKENMLSHLYSSKKKENI 251

QY 54 IKPVFKTIEEKEENKPTFDVSKKKDPQVNHSQLN-----ESHKEDLQR 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 SKENDVLDL-KEEABETEHEEKEEKEETESEISEDEDEDEDEDEDEDEDEDEDE 310

QY 101 EHSQKSDTKDVTATVLDKNNISSKSTNN 131
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 311 EQSNENNDDKDMAE-----QNLISKQNNN 336

RESULT 17
Q9V7J0 PRELIMINARY; PRT; 382 AA.
AC Q9V7J0; Q9C081;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE CG8421 protein (Aspartyl beta-hydroxylase variant 2).
OS ASPH OR CG8421 OR CG18658.
GN Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Rogers J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Watters J.E., Blazek R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry B., Murphy L., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.R., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).

```

---

```

[2] SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J.A., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Ibegwam M., Houck J., Hoskins R.A., Howland D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse C., Li P., Mattel B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bertram B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
[6]
RN [6]
RP SEQUENCE FROM N.A.
RA MEDLINE=20564328; PubMed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved
RT Isoform of Asph Missing the Catalytic Domain Share Exons with
RT Junction."
RL J. Biol. Chem. 275:39543-39554(2000).
DR EMBL; AF003808; AAF58063.2; --
DR EMBL; AF289494; AAG40807.1; --
DR FlyBase; FBgn0034075; Asph.
SQ SEQUENCE 382 AA; 43287 MW; 60E5C03ABBF6E8B CRC64;

Query Match      14.6%; Score 101.5; DB 5; Length 382;
Best Local Similarity 24.5%; Pred. No. 2.8;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 11 EDFTLPVVKGELEKGYQDGEW-----EISFEGKKDAGYVI-----NLSKDTFK 55
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 78 EDLDTPLSESRFSK--VPDGVNDEHRDHDGVDQPSGEALDDHDEHDDHDEDEE 135
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 56 PVFKKIEEKEENKPT-----FDVSKKKDPQVNHSQLNESHKEDLQR 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 PLTELEEELEEEETDEDEPADEEYEDDENNA--GENITAEADAEDEEDND 193

QY 110 TKDVTATVLDKNNISSKST 128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 EGTVEATVEATTEATTEAT 212

RESULT 18
Q9V719 PRELIMINARY; PRT; 556 AA.
AC Q9V719;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

```

4

RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin C.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Hostin N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laeko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulo G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao X., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RN Science 287:2185-2195(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banson J., An H., Baldwin D., Banton J., Beeson K.Y., Buesam D.A.,  
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragaa V., Park S., Patel S., Pfeiffer B., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF289493; AAG40806.1; -;  
DR EMBL; AE003808; AAM70947.1; -;  
DR FlyBase; FBGN0034075; Asph.  
DR InterPro; IPR001440; TPR.  
DR InterPro; IPR006025; Zn MTpeptidase.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
SQ SEQUENCE 785 AA; 89843 MW; 30A8DFCD6836F7F1 CRC64;  
Query Match 14.6%; Score 101.5; DB 5; Length 785;

Best Local Similarity 24.5%; Pred. No. 6;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
Qy 11 EDRFLPVYKGELEKGYQFDGW-----EISFGKKGADAGYVI-----NLSKDTFIK 55  
Db 78 EDLTPLESERFSK--VFDGWVDRHGDHGVQEPSGEALDDHDEHDDHDEDEDEE 135  
Qy 56 PVFKKIEKEEENKPT-----FDVSKKKNDPQVNHSQLNESHKEDLQREHSQKSDS 109  
Db 136 PLTEELSEELSEEEPTEDPADEEVEDEDEENNA--GENITAEADAEDEEDND 193  
Qy 110 TKQVATVLDKNNISSKST 128  
Db 194 EGTVEATVEATTEATTEAT 212  
RESULT 20  
Q9NFV9 PRELIMINARY; PRT; 329 AA.  
ID Q9NFV9 AC Q9NFV9  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Merozoite surface protein 3 (fragment).  
GN MSP3.  
OS Plasmodium falciparum (isolate 7G8).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=57266;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=7G8;  
RX MEDLINE=20416497; PubMed=10960178;  
RA Okeno D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in  
RT Plasmodium reichenowi and Plasmodium falciparum.";  
RL Mol. Biochem. Parasitol. 109:185-188 (2000).  
DR EMBL; AJ252287; CAB85901.1; -;  
FT NON TER 1  
FT NON TER 329  
SQ SEQUENCE 329 AA; 36916 MW; C5B045DB5E21A159 CRC64;  
Query Match 14.4%; Score 100; DB 5; Length 329;  
Best Local Similarity 26.0%; Pred. No. 3.2;  
Matches 38; Conservative 24; Mismatches 44; Indels 40; Gaps 7;  
Qy 1 KEMSSSTVSEEDFLPVYKGELEKGYQFDGWIEISGF--EGKKDAG-----TVINLSKDTF 53  
Db 175 KEASS-----YDYIL-----GWFGGVPHEKKENMLSHLYVSSKDKENI 215  
Qy 54 IKPVFKKIEKEEENKPTFDVSKKKNDPQVNHSQLNESHKEDLQREH-----SQ 105  
Db 216 SKENDVDVLD-KESEAEETEEELKEEETESISEDEEEEEEKEEKEQKEQSQNE 274  
Qy 106 KSDTKQVATVLDKNNISSKSTNN 131  
Db 275 NNDQKDMEA-----QNLISKNNNN 295  
RESULT 21  
Q8IJZ4 PRELIMINARY; PRT; 1130 AA.  
ID Q8IJZ4 AC Q8IJZ4  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF10\_0046.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
Query Match 14.6%; Score 101.5; DB 5; Length 785;



RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
RA Suleston J.E., Craig A., Newbold C., Barrell B.G.;  
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
RL Nature 419:527-531 (2002).  
DR EMBL; AL929355; CAD51734.1; -;  
KW Hydrolyase.  
SQ SEQUENCE 2563 AA; 298805 MW; 3F9613243D26F8F1 CRC64;  
Query Match 14.2%; Score 99; DB 5; Length 2563;  
Best Local Similarity 29.4%; Pred. No. 32;  
Matches 35; Conservative 18; Mismatches 60; Indels 6; Gaps 4;  
Qy 20 GELEKGYQFDG-WEISGFEKGDAGYVILNSKDTFIKPVFKIKBEKEEENKPTFDVSKK 78  
Db 856 GEEKG-NIDGIYILKQKHKKDKMGKEENKFNKSKKEEENKNSNEEIDKNVYLKR 914  
Qy 79 KD---NPNVHNSQLNESHKEDLQRE-HSQKSDSTKDVATVLDKNNISKSSTNNPN 133  
Db 915 KEKHNSNDNEDNDSVLKENVKEETHSSNEQSDSYLKTEKDKNNISVDNEN 973  
RESULT 25  
Q9VQV0 PRELIMINARY; PRT; 157 AA.  
AC Q9VQV0;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE CG31958 protein.  
GN CG31958 OR CG10022.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RC SEQUENCE FROM N.A.  
RX STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.J., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Danke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.D., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Moulton G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.M., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195 (2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banon J.A., An H., Baldwin D., Banon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Seearle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003578; AAF51062.2; -;  
DR HSP; P02593; 1CTR.  
DR FlyBase; FBgn0051958; CG31958.  
DR InterPro; IPR002048; EF-hand.  
DR Pfam; PF00036; ehand; 3.  
DR ProDom; PD000012; EF-hand; 1.  
DR SMART; SM00054; EPH; 3.  
DR PROSITE; PS00018; EF HAND; 1.  
KW Calcium; Calcium-binding.  
SQ SEQUENCE 157 AA; 18263 MW; 543F0480E11D9EC1 CRC64;  
Query Match 14.1%; Score 98; DB 5; Length 157;  
Best Local Similarity 22.7%; Pred. No. 2.1;  
Matches 34; Conservative 29; Mismatches 53; Indels 34; Gaps 5;  
Qy 3 MSTSTVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVILNSKDTFIK 53  
Db 1 MDELSVEEQDLKNIY-SLLDK-----DNEGAITSKELGWMVIRALGROPNESIAKEEF 52  
Qy 54 IKPVFKIEE-KKEEENKPTFDVSKKDN-----POVNHSQLNESHKKE 96  
Db 53 CNVILRKQHTNKEEELRDARFVDEKNNNGYISTELRAVFMALGKLEDELEEMIREY 112  
Qy 97 DLQREHSQKSDSTKDVATVLDKNNISKS 126  
Db 113 DLQDQNHINFEESNNMTTPIILNHTAVK 142  
RESULT 26  
O96229 PRELIMINARY; PRT; 951 AA.  
ID O96229  
AC O96229;





```
Db 194 EGTVEATVEATTEATTEAT 212

RESULT 29
Q98QA1 PRELIMINARY; PRT; 622 AA.
AC Q98QA1;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE LIPOPROTEIN.
GN MYPU_4650.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OC NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RX MEDLINE=21267165; PubMed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA Blanchard A.;
RT "the complete genome sequence of the murine respiratory pathogen
RT Mycoplasma pulmonis.";
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL; AL445564; CAC13638.1; --
DR MyPUList; MYPUL 4650; --
KW Complete proteome.
SQ SEQUENCE 622 AA; 73762 MW; 42BD88930861960D CRC64;

Query Match 13.6%; Score 95; DB 16; Length 622;
Best Local Similarity 25.4%; Pred. No. 15;
Matches 34; Conservative 26; Mismatches 50; Indels 24; Gaps 5;

Qy 8 VSEDFILPVVKGLEKGYQPDGHEISGFEGKKDAGYVINLSKDTFKPVFKKIEKKKE 67
Db 85 LSEDDIIFSLNIENNAKFQDDEF-----VSKDEKFKIKFQEIFSQTE 128
Qy 68 ENKPTFDVSKKKD---NPQVNHSQLNES--HRKEDLQREHSQKSDSTKDVATVLDKN 121
Db 129 Q-KITDNISSKEDSKNNPKDNENNNNSNQKNDQLQNSDKLNDNVQDEKANKNSN 187
Qy 122 -NISSKSTTNPNK 134
Db 188 SNDSKKNENTNK 201

RESULT 30
Q8ILS9 PRELIMINARY; PRT; 3026 AA.
AC Q8ILS9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF14_0165.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014849; AAN36406.1; --
KW Hypothetical protein.
SQ SEQUENCE 3193 AA; 377923 MW; 46B6B21F921C5307 CRC64;

Query Match 13.6%; Score 95; DB 5; Length 3193;
Best Local Similarity 28.3%; Pred. No. 83;
Matches 28; Conservative 21; Mismatches 36; Indels 14; Gaps 3;

Qy 38 GKXDAGYVINLSKDT-----FIKPVFKKIEKKKEE--NKPTFDVSKKKDNPQV 84
Db 1705 GKKSANSINGMIKTSTCNCIEHNDKFKVQYINFLIAQKEKINKEKINEKEKND 1764
Qy 85 NLSQNSHNRKEDLQREHSQKSDSTKDVATVLDKNNI 123
Db 1765 NKXE-EESQKEESQKEESQKEESQKEESKNEENNI 1802

RESULT 32
Q9PPL5 PRELIMINARY; PRT; 312 AA.
AC Q9PPL5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative membrane protein.
GN CU0692C.
```

```
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014818; AAN36777.1; --
KW Hypothetical protein.
SQ SEQUENCE 3026 AA; 357633 MW; 9ECAED915C3C25CE CRC64;

Query Match 13.6%; Score 95; DB 5; Length 3026;
Best Local Similarity 30.2%; Pred. No. 79;
Matches 29; Conservative 15; Mismatches 38; Indels 14; Gaps 4;

Qy 52 TFIKPVFKKIEKKKEENKPTFDVSKKKD---NPQVNHSQLNESHRKEDLQREHSQKS 107
Db 235 TMAVANKKIKNQVEEYKKNVSVINRKDTIQNSDINNTQ-NILHRNDEIEEYKLNEND 293
Qy 108 -DSTKDVATVLDKNNISSKSTT-----NNPNK 134
Db 294 IHNVTKITKEVYSSNSPSSNDDTLTSVESVNNKKNK 329

RESULT 31
Q8IS90 PRELIMINARY; PRT; 3193 AA.
AC Q8IS90;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF11600C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OC NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Nature 419:498-511(2002).
DR EMBL; AE014849; AAN36406.1; --
KW Hypothetical protein.
SQ SEQUENCE 3193 AA; 377923 MW; 46B6B21F921C5307 CRC64;

Query Match 13.6%; Score 95; DB 5; Length 3193;
Best Local Similarity 28.3%; Pred. No. 83;
Matches 28; Conservative 21; Mismatches 36; Indels 14; Gaps 3;

Qy 38 GKXDAGYVINLSKDT-----FIKPVFKKIEKKKEE--NKPTFDVSKKKDNPQV 84
Db 1705 GKKSANSINGMIKTSTCNCIEHNDKFKVQYINFLIAQKEKINKEKINEKEKND 1764
Qy 85 NLSQNSHNRKEDLQREHSQKSDSTKDVATVLDKNNI 123
Db 1765 NKXE-EESQKEESQKEESQKEESQKEESKNEENNI 1802

RESULT 32
Q9PPL5 PRELIMINARY; PRT; 312 AA.
AC Q9PPL5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative membrane protein.
GN CU0692C.
```



```

RESULT 34
002124 PRELIMINARY; PRT; 210 AA.
AC O02124;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DI 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE E03H12.5 protein.
GN GN E03H12.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2; MEDLINE=94150718; PubMed=7906398;
RX Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Bonfield J., Burton S., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jfer M., Johnston L., Jones M., Kershaw J., Kirsten J., Laister N., Latreille P., James M., Kershaw J., Kirsten J., Laister N., Latreille P., Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M., Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R., Smalton N., Smith A., Sonhammer E., Staden R., Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.; "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."; Nature 368:32-38 (1994).
RL RL Nature 368:32-38 (1994).
RP [2]
RC SEQUENCE FROM N.A.
RX STRAIN=Bristol N2;
RA Nelson, J., Wohldmann, P., Sansone J.;
RT "The sequence of C. elegans cosmid E03H12."; Submitted (JUN-1997) to the EMBL/GenBank/DDJB databases.
DR EMBL; AF000299; AAC47980.1; WormPep; E03H12.5; CE09144.
SQ SEQUENCE 210 AA; 23703 MW; 641B64BCF7AC983B CRC64;

Query Match          13.5%; Score 94; DB 5; Length 210;
Best Local Similarity 28.6%; Pred.No. 5.8;
Matches 28; Conservative 18; Mismatches 52; Indels 0; Gaps 0

QY      37 EGKDGAGVIVNLSDTFIKPVFKIEKEEENKTPTFDVSKKKDNPQNHNHSQLNESHKRKKD
Db       69 EGEEKGDKGEKKGDKGGKEEKDDKEDGDCKDEDKDEKDKEDKDEKDKKDADEKDK
QY      97 DLQREHSQSXDSTKVTVATVLQKNNISSSKTTNNPNK 134
Db       129 EKDKDKDEKDEKDEKDEKESKSKSSEKKSSESSEKSKS 166

RESULT 35
017595 PRELIMINARY; PRT; 535 AA.
AC Q17595;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DI 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 60.3 kDa protein.
GN CO2H7.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2; MEDLINE=99069613; PubMed=9851916;
RX RA None;
RT "genomic sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium.";
```



```

(2)
RN SEQUENCE FROM N.A.
RP SPECIES=E.faecalis; STRAIN=RE25; PLASMID=PRE25;
RX PubMed=11735367;
RT Schwarz F.V., Perreten V., Teuber M.;
RT "Sequence of the 50-kb conjugative multiresistance plasmid PRE25 from
RT Enterococcus faecalis RE25.";
RL Plasmid 46:170-187(2001).
DR EMBL; L39769; AAA99471.1; -
DR EMBL; X92945; CAC29184.1; -
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 450 AA; 52755 MW; F3CA5C827CFF30CB CRC64;

Query Match 13.4%; Score 93; DB 2; Length 450;
Best Local Similarity 26.0%; Pred. No. 15;
Matches 39; Conservative 28; Mismatches 57; Indels 26; Gaps 7;

QY 9 SEEDFLPVYK-GELEK-----GYQFDGWEISGEGKKDAGVYNLSKDTFIKPVF 58
DB 302 SDKKIQSLYKQGETEKIEALDPLGSDYLADPKKILAYDNSDIEYLVQSTDTTII--VE 359
QY 59 KKEIEKEEENKPT-----FDVSKKK---DNPQVNSQLNESHKEDLQR--BEHSQK 106
DB 360 ALIDSKVKEKDIPYKVNQVSIQKQLAIDPKRQISMIDLLIENNSELENMYKDSNLN 419
QY 107 SDSTKDVATVLDKNN--ISSKSTNNPNK 134
DB 420 EDLKKKQTNLDLENNNTLLSEKIETNAEK 449

RESULT 39
QBMN05 PRELIMINARY; PRT; 455 AA.
AC Q8MN05;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
(1)
RN SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Fachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC117177; AAM43671.1; -
KW Hypothetical protein.
SQ SEQUENCE 455 AA; 51270 MW; D8B00463D1FEA2A4 CRC64;

Query Match 13.4%; Score 93; DB 5; Length 455;
Best Local Similarity 32.9%; Pred. No. 16;
Matches 24; Conservative 10; Mismatches 33; Indels 6; Gaps 1;

QY 62 EEKKEEENKPTFDVSKKDNPNQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKN 121
DB 201 EEDEEVKVLMDIDTDTTKSKQKQKE-----NEDDEEEEGDIIKSTKDLNLLVNNN 254
QY 122 NISSKSTNNPNK 134
DB 255 NSSITSSLNYPNK 267

RESULT 40
Q8IDA0 PRELIMINARY; PRT; 3127 AA.
ID Q8IDA0;
AC Q8IDA0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

```

```

DE Guanylyl cyclase (EC 4.6.1.2).
GN GC-BETA.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN (1)
RP SEQUENCE FROM N.A.
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
RA Herriman M., Pain A., Hall N., Arkin R., Chillingworth C., Doggett J.,
RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrall B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844509; CAD52725.1; -
KW Lyase.
SQ SEQUENCE 3127 AA; 370597 MW; F0375C72B9007560 CRC64;

Query Match 13.4%; Score 93; DB 5; Length 3127;
Best Local Similarity 24.7%; Pred. No. 1.2e+02;
Matches 45; Conservative 29; Mismatches 58; Indels 50; Gaps 7;

QY 1 KEMSTIVSEEDFILPVYKGELEKGYQFDGWEISG---FEGKKDAGVYNLSKDTFIKPV 57
DB 2000 KKEINTIVSDDMFTSPVNIKEYNNEQERKKEIVGNLSYDKTKKIFPFKFTKEGRKK- 2058
QY 58 FKKIEEKEEENK-----PTF-----DVSKKKNQOV 84
DB 2059 -KKIEKKEKKEKKNNNFLYNDYSSSPKYGDNENNFKYIRERKDKQKKEHDHNF 2117
QY 85 NHSQL-----NESHK-----EDLQREHSQKSDSTKD-VTATVLDKNINISSKSTNNP 132
DB 2118 NFSKFLHYNPMKKNKKNKKNKNNRNEYNPNTSSSKDGVSYNFLSDLSFSSDNEYSSD 2177
QY 133 NK 134
DB 2178 NE 2179

Search completed: February 10, 2004, 10:57:04
Job time : 24.725 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:57:15 ; Search time 27.9011 Seconds  
(without alignments)  
1230.730 Million cell updates/sec

Title: us-10-067-385-8\_copy\_610\_773

Perfect score: 848

Sequence: 1 TTWKFTLNKDTGVSRLKPV...ATVLDKNNISSKSTNNPNK 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT NEW PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	848	100.0	773	14	US-10-067-385-8
2	848	100.0	2119	12	US-09-769-744A-28
3	615	72.5	117	9	US-09-765-272-68
4	106	12.5	665	11	US-09-820-843A-107
5	96.5	11.4	2060	12	US-10-381-596A-2
6	93.5	11.0	654	12	US-10-172-502-10
7	92	10.8	1875	12	US-10-369-493-22285
8	90.5	10.7	635	12	US-10-032-585-7046
9	90.5	10.7	743	15	US-10-171-311-188
10	90.5	10.7	758	9	US-09-925-299-859
11	90.5	10.7	758	11	US-09-925-299-859
12	89.5	10.6	402	10	US-09-827-664-4
13	89.5	10.6	511	12	US-10-289-762-509
14	89.5	10.6	516	12	US-10-032-585-7407
15	89.5	10.6	529	10	US-09-827-664-2

16	88	10.4	225	12	US-10-032-585-7829	Sequence 7829, Ap
17	88	10.4	1298	12	US-10-144-194A-106	Sequence 106, App
18	88	10.4	3051	12	US-10-144-194A-62	Sequence 62, Appl
19	87.5	10.3	655	12	US-10-369-493-2016	Sequence 2016, Ap
20	87.5	10.3	1349	9	US-09-815-242-5898	Sequence 5898, Ap
21	87.5	10.3	1349	9	US-09-815-242-13137	Sequence 13137, A
22	87.5	10.3	1363	12	US-10-032-585-7800	Sequence 7800, Ap
23	87.5	10.3	1702	11	US-09-839-996-5	Sequence 5, Appli
24	87.5	10.3	1702	15	US-10-080-505-5	Sequence 16456, A
25	87	10.3	323	12	US-10-369-493-16456	Sequence 21889, A
26	87	10.3	778	12	US-10-369-493-21889	Sequence 5013, Ap
27	87	10.3	6642	12	US-10-369-493-5013	Sequence 14032, A
28	86.5	10.2	443	9	US-09-815-242-14032	Sequence 18, Appl
29	86.5	10.2	732	9	US-09-874-923-18	Sequence 18, Appl
30	86.5	10.2	732	10	US-09-991-496-18	Sequence 5292, Ap
31	86.5	10.2	806	12	US-10-369-493-5292	Sequence 10918, A
32	86.5	10.2	891	12	US-10-369-493-10918	Sequence 32082, A
33	85.5	10.1	204	12	US-10-029-386-32082	Sequence 2272, Ap
34	85.5	10.1	734	12	US-10-104-047-2272	Sequence 22302, A
35	85.5	10.1	752	12	US-10-369-493-22302	Sequence 7212, Ap
36	85.5	10.1	884	12	US-10-032-585-7212	Sequence 22375, A
37	85.5	10.1	1007	12	US-10-369-493-22375	Sequence 5200, Ap
38	85	10.0	208	8	US-08-781-986A-5200	Sequence 1313, Ap
39	85	10.0	259	10	US-09-764-864-1313	Sequence 5838, A
40	85	10.0	636	9	US-09-815-242-5838	Sequence 13008, A
41	85	10.0	636	9	US-09-815-242-13008	Sequence 3899, Ap
42	84.5	10.0	390	10	US-09-738-626-3899	Sequence 7146, Ap
43	84.5	10.0	434	12	US-10-032-585-7146	Sequence 5578, Ap
44	84	9.9	932	9	US-09-815-242-5578	Sequence 1438, A
45	84	9.9	932	9	US-09-815-242-12438	

ALIGNMENTS

RESULT 1  
US-10-067-385-8  
; Sequence 8, Application US/10067385  
; Publication No. US20020110562A1  
; GENERAL INFORMATION:  
; APPLICANT: Adamou, John  
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines  
; FILE REFERENCE: 469201-589  
; CURRENT APPLICATION NUMBER: US/10/067,385  
; CURRENT FILING DATE: 2002-02-05  
; PRIOR APPLICATION NUMBER: US/09/590,991  
; PRIOR FILING DATE: 2000-06-09  
; PRIOR APPLICATION NUMBER: US/60/138,453  
; PRIOR FILING DATE: 1999-06-10  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 773  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-067-385-8

Query Match	100.0%	Score	848	DB	14	Length	773
Best Local Similarity	100.0%	Pred. No.	2.9e-68				
Matches	164	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	TTWKFTLNKDTGVSRLKPVHVTVTIQNGKMSSTIVSBE	DFILPVYKGLKGYQFDG	60			
Db	610	TTWKFTLNKDTGVSRLKPVHVTVTIQNGKMSSTIVSBE	DFILPVYKGLKGYQFDG	669			
Qy	61	WEISGFGKGDAGVYINLSKDTFIKPVFKIEEKEEENKPT	FDVSKKONPQVNHSQLN	120			
Db	670	WEISGFGKGDAGVYINLSKDTFIKPVFKIEEKEEENKPT	FDVSKKONPQVNHSQLN	729			
Qy	121	ESHRKEDLQREHSQKSDSKDVATATVLDKNNISSKSTNN	PNK 164				
Db	730	ESHRKEDLQREHSQKSDSKDVATATVLDKNNISSKSTNN	PNK 773				

```
RESULT 2
US-09-769-744A-28
; Sequence 28, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28

Query Match          100.0%; Score 848; DB 12; Length 2119;
Best Local Similarity 100.0%; Pred. No. 1e-67; Indels 0; Gaps 0;
Matches 164; Conservative 0; Mismatches 0;

QY 1 TTVKEFILNDKGEVSELKPHRVTVTIQNGKEMSSIVSEEDFILPVYKGELEKGYQFDG 60
DB 1922 TTVKEFILNDKGEVSELKPHRVTVTIQNGKEMSSIVSEEDFILPVYKGELEKGYQFDG 1981

QY 61 WEISGEGKDGAGYVNLKDTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLN 120
DB 1982 WEISGEGKDGAGYVNLKDTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLN 2041

QY 121 ESHRKEDLOREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 164
DB 2042 ESHRKEDLOREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 2085

RESULT 3
US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
```

```
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 117 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68

Query Match          72.5%; Score 615; DB 9; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.2e-48; Indels 0; Gaps 0;
Matches 117; Conservative 0; Mismatches 0;

QY 48 YKGELEKGYQFDGWEISGEGKDGAGYVNLKDTFIKPVFKKIEKKEENKPTFDVSK 107
DB 1 YKGELEKGYQFDGWEISGEGKDGAGYVNLKDTFIKPVFKKIEKKEENKPTFDVSK 60

QY 108 KDNQVNHSQLNESHKEDLOREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 164
DB 61 KDNQVNHSQLNESHKEDLOREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117

RESULT 4
US-09-820-843A-107
; Sequence 107, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 107
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc feature
; OTHER INFORMATION: gi|3845248
US-09-820-843A-107

Query Match          12.5%; Score 106; DB 11; Length 665;
Best Local Similarity 24.3%; Pred. No. 0.33; Indels 56; Gaps 10;
Matches 42; Conservative 35; Mismatches 40;

QY 10 KDTGEVSELKPHRVTVTIQNGKEMSSIVSEEDFILPVYKGELEKGYQFDGWEI--SGF 66
DB 127 EKNKINKSDLHRQNELNLSQSK-----NEQDI-----NKNEKGKQ----DISNSNA 169

QY 67 EGKDGAGYVNLKDTFIKPVFKKIEKKE-----BENKPTFD-----VSKKKNP 112
DB 170 ENKDK-----VKEGVKELEKKEEKEIKSDHVKVEENKSDHDKHVEENKSDDH 217

QY 113 QVNHSQLNESHKEDLOREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 164
DB 218 KVEENKSDHDKHKEEVKVEEHEDEE-----DKKKEKSENKKNKDNK 261

RESULT 5
US-10-381-596A-2
```

Query Match 10.7%; Score 90.5; DB 12; Length 635;  
Best Local Similarity 22.2%; Pred. No. 7.8;  
Matches 38; Conservative 40; Mismatches 70; Indels 23; Gaps 8;





Db 502 KEFVRHKEKIQAKEA-VKENLKFSDFSVKSTFRHFKDTTKNIFDE 547

## RESULT 12

US-09-827-664-4

; Sequence 4, Application US/09827664

; Patent No. US20020150956A1

; GENERAL INFORMATION:

; APPLICANT: Lawlor, Elizabeth J.

; TITLE OF INVENTION: No. US20020150956A1el tig

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESS: Dechert Price & Rhoads

; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

; CITY: Philadelphia

; STATE: PA

; COUNTRY: US

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; FILING DATE: 06-Apr-2001

; CLASSIFICATION: <unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/464,483

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Falk, Stephen T

; REGISTRATION NUMBER: 36,795

; REFERENCE/DOCKET NUMBER: GMI0085

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 215-994-2488

; TELEFAX: 215-994-2222

; TELETYPE: <unknown>

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 402 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-09-827-664-4

Query Match 10.6%; Score 89.5; DB 10; Length 402;  
Best Local Similarity 22.1%; Pred. No. 5, 4;  
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

Qy 11 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK-----54

Db 53 DETDIKPAQEVSVTQIEKGDFFTEATVVEPEVKLGDKYGLIEKQETLSDELQE 112

Qy 55 -----GYQFDGWELSG--FEGKKDAGVYVNLKDTTFKP 86

Db 113 AIDHSLGHLAEMVVKEDGVWENGDTVNI DFGS-SVDGEFEFGQAGYDLIEIGSGSFTI-P 170

Qy 87 VFK-----KTEEKE-----EE--NKPTFDVS-----KKKDNPNVHNSQLNE- 121

Db 171 GFEEQLEGMKVDDEKDVVVTFPEYHAEELAGKEATFKYKNEIKFKEVPBELTBEIANEL 230

Qy 122 -----SHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTN 160

Db 231 DAEANTVDEYKENLKRKLAEQKTAENV-----EKEEAITKATDN 271

## RESULT 13

US-10-289-762-509

; Sequence 509, Application US/10289762

; Publication No. US20040006218A1

; GENERAL INFORMATION:

; APPLICANT: Griffiths, R.

; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments

; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve

; FILE REFERENCE: 9710-003-999

; CURRENT APPLICATION NUMBER: US/10/289,762

; CURRENT FILING DATE: 2003-03-27

; NUMBER OF SEQ ID NOS: 6849

; SEQ ID NO 509

; LENGTH: 511

; TYPE: PRT

; ORGANISM: Chlamydia pneumoniae

US-10-289-762-509

Query Match 10.6%; Score 89.5; DB 12; Length 511;

Best Local Similarity 24.5%; Pred. No. 7, 3;

Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;

Qy 84 IKPVFKKIEEKEEENKPTFD-----VSKKDNPNVHNSQLNESHRK 125

Db 95 VKGVFKTTPQARPEVSSPRLPESHVHGQRLPGLSGFRDRIOKRSENPADLGKKRSYS 154

Qy 126 EDLQREHSQKSDTKDVTATVLDKNNISSKSTT 159

Db 155 GDLDKRVGHSDNEDSTEDSR---EGGEPPSSKSS 185

## RESULT 14

US-10-032-585-7407

; Sequence 7407, Application US/10032585

; Publication No. US20030180953A1

; GENERAL INFORMATION:

; APPLICANT: Terry, Roemer D.

; APPLICANT: Bo, Jiang

; APPLICANT: Charles, Boone

; APPLICANT: Howard, Bussey

; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery

; FILE REFERENCE: 10182-005-999

; CURRENT APPLICATION NUMBER: US/10/032,585

; CURRENT FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 8000

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 7407

; LENGTH: 516

; TYPE: PRT

; ORGANISM: Candida albicans

US-10-032-585-7407

Query Match 10.6%; Score 89.5; DB 12; Length 516;

Best Local Similarity 23.3%; Pred. No. 7, 4;

Matches 37; Conservative 27; Mismatches 58; Indels 37; Gaps 6;

Qy 10 KDTGEVSELKPHRVTVT--IQNGKMSSTIVSEEDFILPVYKGELEKGYQF-----58

Db 369 KNGKRISRYLANKCSIASRIDYSEPTAFGE-----ILKKQVEDRLKFDYDTGSAPMK 422

Qy 59 -----DGWEISGFEKGKQAGVYVNLKDTTFIKPVFKIEEKE-----BENKPTF 103

Db 423 NSDAIKAAALNGQDLAGABEQKQDMV---SDEQVKKEKKEKKEKKEKKEKKEKKEK 479

Qy 104 DVSKKDNPNVHNSQLNESHRKEDLQREHSQKSDTKD 142

Db 480 DKKEKKKKKKKKKK--DKKRKSDGGETPKKKKKKKSKD 516

## RESULT 15

US-09-827-664-2

; Sequence 2, Application US/09827664

; Patent No. US20020150956A1

; GENERAL INFORMATION:

; APPLICANT: Lawlor, Elizabeth J.

; TITLE OF INVENTION: No. US20020150956A1el tig

```
;
;
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/827,664
; FILING DATE: 06-Apr-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/464,483
; FILING DATE: <Unknown>
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Falk, Stephen T
; REGISTRATION NUMBER: 36,795
; REFERENCE/DOCKET NUMBER: GM10085
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-994-2488
; TELEFAX: 215-994-2222
;
; INFORMATION FOR SEQ ID NO: 2:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-827-664-2
```

```
Query Match 10.6%; Score 89.5; DB 10; Length 529;
Best Local Similarity 22.1%; Pred. No. 7.6;
Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;

QY 11 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 54
Db 81 DETDIRKPAQVSVTVQIEKGDIFFEATVVEPVKLDGYKGLTEKQETELSDDELOE 140
QY 55 -----GYQFDGWEISG--FEGKQDAGYVINLSKDTFIKP 86
Db 141 AIDHSILGLAEMVVKEDGVVNGDVTNIDFSG-SVDGEFEGGQAGYDLEIGSGSFI-P 198
QY 87 VPK-----KIBEKE-----EE--NKPTFDVS-----KKQNPQVNHSQLNE- 121
Db 199 GFEEQLGKMKVDEKDVVTFPPEVHAEELAGKEATFKTVNKEIKPKVEPFLTDEIANEL 258
QY 122 -----SHRKEDLQREHSQKSDSTKYDTATVLDKNNISSKSTTN 160
Db 259 DAENATVDEYKENLRKRLAEQKATDAENV-----EKEEAITRATDN 299
```

```
RESULT 16
US-10-032-585-7829
; Sequence 7829, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
;
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 7829
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7829

Query Match 10.4%; Score 88; DB 12; Length 225;
Best Local Similarity 28.0%; Pred. No. 3.5;
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;

QY 56 YQFDGWEISGFEKGDAGVINLSKDTFIKPVFKK-----IEEKKEENKPTFDVSKKK-- 109
Db 80 YDDDDDEFEFESSNGAAKELNLSQAKWKQRDLIEIEREKLNSKKKEIEIEKAKS 139
QY 110 --DNPQVNHSQLNESHKEDLQREH--SQKSDSTKYDTATVLDKNN 152
Db 140 TIDDFYENYNKRDNHQKEILLSEKQKFIKRDDFLK--RGTILDRVN 184

RESULT 17
US-10-144-194A-106
; Sequence 106, Application US/10144194A
; Publication No. US20030215809A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies Inc
; TITLE OF INVENTION: Regulated Breast Cancer Genes
; FILE REFERENCE: 3U 103 R1
; CURRENT APPLICATION NUMBER: US/10/144,194A
; CURRENT FILING DATE: 2002-06-12
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 106
; LENGTH: 1298
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-194A-106

Query Match 10.4%; Score 88; DB 12; Length 1298;
Best Local Similarity 25.9%; Pred. No. 33;
Matches 45; Conservative 23; Mismatches 60; Indels 46; Gaps 8;

QY 10 KDTGVSSELKPHRVTVTQNGKMSSTIVSEEDFILPVYKGLKGYQFDGWEISGFEK 69
Db 944 KPSG-IASLQREPLVNLVNSLGVNFTNEIHSESYNKGIEISSRKDNAAEISGHSVE 1002
QY 70 KDAGYVINLSKDTFIKPVFKKIEEKKEENKPTFDVSKKK-----DNPQVNHSQL 118
Db 1003 ADP-----KEVEE--EERHMPK--RKRKQHYLSSEDEPDNDPDVLDNR 1041
QY 119 LNESHR-----KEDLQ--BEHSQKSDSTKYDTATVLDKNNISSKSTT 159
Db 1042 IETAQRCQPETEPHPTKEENSRDLBELPKTSSETNSTTSRVNMEKDEYSSSETT 1095

RESULT 18
US-10-144-194A-62
; Sequence 62, Application US/10144194A
; Publication No. US20030215809A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies Inc
; TITLE OF INVENTION: Regulated Breast Cancer Genes
; FILE REFERENCE: 3U 103 R1
; CURRENT APPLICATION NUMBER: US/10/144,194A
; CURRENT FILING DATE: 2002-06-12
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 62
; LENGTH: 3051
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-194A-62

Query Match 10.4%; Score 88; DB 12; Length 3051;
```

Best Local Similarity 25.9%; Pred. No. 96;  
Matches 45; Conservative 23; Mismatches 60; Indels 46; Gaps 8;  
  
Qy 10 KOTGEVSELPKRVTVTTIQNGKEMSSIVSEDFILPVYKGELEKGYQDFDGEWISGPEGK 69  
Db 2697 KPQG-IAELQREPLDLVNSLNVNSGFTNEEIIHSESYNKGESISGRKNAEALSHSVE 2755  
  
Qy 70 KDAGYVNLSDTFTKPVFKKIEEKEENKPTFDVSKK-----DNPQVNHQS 118  
Db 2756 ADP-----KEVEE--EERHMPK---RKRKHLYLSSEDEPDNDVLDLSR 2794  
  
Qy 119 LNEHSHR-----KEDLQR--EHSQKSDSTKDVATVL-DKNNISKSTT 159  
Db 2795 IETAQRCQPEPHDTKEENSRDLLELPKTSSETNSTSRVMEKDEYSSETT 2848  
  
RESULT 19  
US-10-369-493-2016  
; Sequence 2016, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 2016  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-2016

Query Match 10.3%; Score 87.5; DB 12; Length 655;  
Best Local Similarity 20.5%; Pred. No. 15;  
Matches 34; Conservative 36; Mismatches 57; Indels 39; Gaps 7;  
  
Qy 6 FILNKOTGEVSELPKRVTVTTIQNGKEMSSIVSEDFILPVYKGELEKGYQF 58  
Db 209 YIINKRTNSIK-----RSVSRTLRKGK-----TDSILPVYQSELKPFPRPSDDDDYK 255  
  
Qy 59 DGWEISGF--EGKKGAGYVNLSDTFTKPVFKKIEEKEE-----ENKPTFDVSK 107  
Db 256 TNIEDNKVREGR-----VHVSKESTADSTQKLGKEQKVIQSHLRHHDNNSTFRPHR 309  
  
Qy 108 KDNQPV--NHSQNLNESHKEDLQREHSQKSDSTKQVATVLDKN 151  
Db 310 LAPAPATKNHDSKTKWKHEDLLELKNNDSDSNEIIMKMTVAIDVN 355  
  
RESULT 20  
US-09-815-242-5898  
; Sequence 5898, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELTRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSEQ for Windows Version 4.0  
SEQ ID NO 5898  
LENGTH: 1349  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5898  
  
Query Match 10.3%; Score 87.5; DB 9; Length 1349;  
Best Local Similarity 25.0%; Pred. No. 38;  
Matches 41; Conservative 23; Mismatches 73; Indels 27; Gaps 5;  
  
Qy 9 NKDTGEVSELPKRVTVTTIQNGKEMSSIVSEDFILPVYKGELEKGYQDFDGEWISGF- 66  
Db 804 NKDGKQDSTKSGISGVTVTLKN-----ENGEVLQTTKDKGKYQFTGLENGTYK 853  
  
Qy 67 -EGKKGAGYVNLSDTFTKPVFKKIEEKEENKPTFDVSKKDNQVNHQSOLNESHK 125  
Db 854 VEFETPSGYPT-----QVSGTDEGDSNCTSTTGVIKDKNDTIDSGFYKPTVNL 905  
  
Qy 126 EDLQREHSQKSDSTKD-----VTATVLDKNNISKSTNNPN 163  
Db 906 GDYVWEDTNKNGVQDKDEKGISGVTVTLKDENDKVLKTVTTDEN 949  
  
RESULT 21  
US-09-815-242-13137  
; Sequence 13137, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELTRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16





APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14032  
LENGTH: 443  
TYPE: PRT  
ORGANISM: Salmonella typhi  
US-09-815-242-14032

Query Match 10.2%; Score 86.5; DB 9; Length 443;  
Best Local Similarity 23.5%; Pred. No. 11; Mismatches 62; Indels 57; Gaps 8;  
Matches 43; Conservative 21; Mismatches 62; Indels 57; Gaps 8;  
QY 16 SELKPHRVTVTTONKEMSTIVSE-----EDFILPVYKGELEKGYQ 57  
Db 13 SRFTSLRTHITMNLTKNTPVSELITLGSWGLENLARMKQDIIIFAILQHAQSGED 72  
QY 58 F-----DQWEISGEGKDGAGYVNLKSDTFIKP-----VFKKIEK 94  
Db 73 IFDGVLEILQDGF---GFLRSADSSYLAG-PDDIYVSPSQRIRFLRTGDTISGIRPP 128  
QY 95 KEE-----NKPTFDVSKKKNPQVNHSQLNESHKEDIQREHSQKSDSTKDVAT 146  
Db 129 KEGERYFALLKVNVDYKPNARN-KILFENLTPLHANSRLRME---RGNSTGSDLTAR 184  
QY 147 VLD 149  
Db 185 VLD 187

RESULT 29  
US-09-874-923-18  
Sequence 18, Application US/09874923  
Patent No. US20020081320A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Coler, Rhea  
APPLICANT: Probst, Peter  
APPLICANT: Brannon, Mark  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
FILE REFERENCE: 210121.420C8

CURRENT APPLICATION NUMBER: US/09/874,923  
CURRENT FILING DATE: 2001-06-04  
NUMBER OF SEQ ID NOS: 122  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 732  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-874-923-18

Query Match 10.2%; Score 86.5; DB 9; Length 732;  
Best Local Similarity 23.6%; Pred. No. 22;  
Matches 39; Conservative 29; Mismatches 58; Indels 39; Gaps 8;  
QY 6 FILNKDTGEVSELKPHRVTVTTONKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISG 65  
Db 170 FTVRTDTGE---PMGRGTVKVLHLKEDQTEYLEERRI-----KEIVKHSQFIGYPITL 220  
QY 66 F-EGKKDAGYVNLKSDTFIKPVFKKIEEKEENKPTFDVSKKKNPQVNHSQLNESHHR 124  
Db 221 FVEKERDK-----EVSDD-----EABEKEDKEEKEEKESEDKPEI----- 258  
QY 125 KEDLQREHSQKSD-----STKDVATVLDKNNISSKST--TNNPN 163  
Db 259 -EDVGSDEDEKKGDKGKKKKKIKKIKYIDKSELNKTPIWTRNPD 302

RESULT 30  
US-09-991-496-18  
Sequence 18, Application US/09991496  
Patent No. US20020169285A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Campos-Neto, Antonio  
APPLICANT: Webb, John R.  
APPLICANT: Dillon, Davin C.  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Bhatia, Ajay  
APPLICANT: Coler, Rhea  
APPLICANT: Probst, Peter  
APPLICANT: Brannon, Mark  
TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE  
FILE REFERENCE: 210121.420C9  
CURRENT APPLICATION NUMBER: US/09/991,496  
CURRENT FILING DATE: 2001-11-20  
NUMBER OF SEQ ID NOS: 137  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 732  
TYPE: PRT  
ORGANISM: Homo sapien  
US-09-991-496-18

Query Match 10.2%; Score 86.5; DB 10; Length 732;  
Best Local Similarity 23.6%; Pred. No. 22;  
Matches 39; Conservative 29; Mismatches 58; Indels 39; Gaps 8;  
QY 6 FILNKDTGEVSELKPHRVTVTTONKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISG 65  
Db 170 FTVRTDTGE---PMGRGTVKVLHLKEDQTEYLEERRI-----KEIVKHSQFIGYPITL 220  
QY 66 F-EGKKDAGYVNLKSDTFIKPVFKKIEEKEENKPTFDVSKKKNPQVNHSQLNESHHR 124  
Db 221 FVEKERDK-----EVSDD-----EABEKEDKEEKEEKESEDKPEI----- 258  
QY 125 KEDLQREHSQKSD-----STKDVATVLDKNNISSKST--TNNPN 163  
Db 259 -EDVGSDEDEKKGDKGKKKKKIKKIKYIDKSELNKTPIWTRNPD 302  
RESULT 31  
US-10-369-493-5292

TITLE OF INVENTION: NO. US2003023692A1el full length cDNA  
 FILE REFERENCE: HI-A0105  
 CURRENT APPLICATION NUMBER: US/10/104,047  
 CURRENT FILING DATE: 2002-03-25  
 PRIOR APPLICATION NUMBER:  
 PRIOR FILING DATE:  
 NUMBER OF SEQ ID NOS: 4096  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 2272  
 LENGTH: 734





```

; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781.986A
; FILING DATE:
;
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: P8248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 5200:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 208 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-781-986A-5200
;
; Query Match 10.0%; Score 85; DB 8; Length 208;
; Best Local Similarity 21.7%; Pred. No. 6;
; Matches 38; Conservative 22; Mismatches 89; Indels 26; Gaps 5;
;
; Qy 9 NKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWELSGFEG 68
; Db 27 NSSKNDTNKTKTQTDNTQSTNTEKQMTTPQBAEDIVRNDYKARGVNEYQTILNYKTNLRS 86
; Qy 69 KDAQGVNLSKDTFIKP-----VFKIEEKKEENKPTFDVSKK---K 109
; Db 87 NEHEYVEHLVRDAVGTFLKCAIVNRNGNIINIFDDMSKDEE-----FEAFKRSFK 142
; Qy 110 DNPQVNSQLNESHKEDLQREHHSQKSDSTKDVATVLDKNNISSKSTTNPNK 164
; Db 143 YNPGMNHDETDG-ESEDIQHIDNNKAIQNDIPQKVDDKN--DKNAVNKEEK 194
;
; RESULT 39
; US-09-764-864-1313
; Sequence 1313, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1313
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (251)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
;
; ADDRESS: (255)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (259)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-09-764-864-1313
;
; Query Match 10.0%; Score 85; DB 10; Length 259;
; Best Local Similarity 23.8%; Pred. No. 7.9;
; Matches 36; Conservative 30; Mismatches 51; Indels 34; Gaps 8;
;
; Qy 17 ELKPHRVTVTIQ-----NGKEM--SSTIVSEEDFILPVYKGELEKGYQ-FDGEWELSGF 66
; Db 93 EMKNDRIKVSLSMKVNVNQGTGKOLDPNNVIIQEE-----RRRSFQDYTCOKIT-L 143
; Qy 67 EGK-----KDAGYVINLSKDTFIKPVFKJ-----EKKKEENKPTFDVSKKDNQVNV 115
; Db 144 EAVLNTTKCKGCGKGFPAKCFQPGGTGKYSLIPDEEEKEEAKSAEFE-----KPDPTRN 199
; Qy 116 HSQNLNESHKEDLQREHHSQKSDSTKDVAT 146
; Db 200 PSRRKKEKKKKKXDRKSDSDSDSDSDT 230
;
; RESULT 40
; US-09-815-242-5838
; Sequence 5838, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR FILING DATE: 60/191,078
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5838
; LENGTH: 636
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-815-242-5838
;
; Query Match 10.0%; Score 85; DB 9; Length 636;
; Best Local Similarity 22.0%; Pred. No. 25;
; Matches 35; Conservative 35; Mismatches 61; Indels 28; Gaps 7;
;
; Qy 7 ILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFDGWELSGF 66
; Db 27 IVTKDYGSKSQVNSAGSKNGTILDSRYLSALYLYEDI--IYAIGLTNKYE-----GD 78
; Qy 67 EGKXKDAQGVNLSKDTFIKPVFKK-----IEBKKEE-ENKPTFDVSKKDNQVNVHSQLN- 120
```

Db 79 NIYKEA-----KDLLEKVLREDQYLLERKKSQYEDYKQWYANYKENPRTDLKMANF 131  
QY 121 ESHRKEDLOREHHSQKSDTKDV-----TATVLDKNN 152  
Db 132 HKYNLEELSMKEYNELQDALKRALDDFHREVKDIDKNS 170

Search completed: February 10, 2004, 11:23:56  
Job time : 28.9011 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:45 ; Search time 11.1197 Seconds  
(without alignments)  
1331.870 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799

Sequence: 1 DTGEVSELKPHRVTVTIQNG.....ATVLDKNNISKSTNNPNK 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	799	100.0	2140	2 F95074	serine proteinase,
2	796	99.6	2144	2 A97942	metalloproteinase,
3	110	13.8	558	2 T18467	hypothetical prote
4	108.5	13.6	1038	2 JCS497	claustrin - chicke
5	105	13.1	665	2 B71609	hypothetical prote
6	103.5	13.0	325	2 T18283	hypothetical prote
7	100	12.5	312	2 G81339	probable membrane
8	99	12.4	211	2 T25911	hypothetical prote
9	97.5	12.2	614	2 A84152	hypothetical prote
10	97.5	12.2	1345	2 S46817	hypothetical prote
11	97	12.1	385	2 T20410	hypothetical prote
12	97	12.1	988	2 T14188	hypothetical prote
13	96.5	12.1	540	2 D86432	hypothetical prote
14	96	12.0	535	2 T37189	hypothetical prote
15	96	12.0	2500	2 G71609	hypothetical prote
16	95.5	12.0	644	2 T47835	hypothetical prote
17	95	11.9	348	2 I37271	cylicin II - human
18	95	11.9	622	2 A90570	lipoprotein (impor
19	95	11.9	1397	2 T10466	DNA topoisomerase
20	94.5	11.8	3724	2 T18427	hypothetical prote
21	94	11.8	210	2 T28771	hypothetical prote
22	93.5	11.7	456	2 T05612	hypothetical prote
23	93	11.6	219	2 B72291	hypothetical prote
24	91	11.4	253	2 T32879	hypothetical prote
25	90.5	11.3	629	2 G36542	hypothetical prote
26	90.5	11.3	867	2 T28391	ORF MSV230 hypothe
27	90.5	11.3	867	2 T27136	hypothetical prote
28	90.5	11.3	871	2 T27135	hypothetical prote
29	90.5	11.3	1332	2 S41552	probable transcrip

30	90.5	11.3	2401	2 T28676	rhoptyr protein -
31	90	11.3	645	2 E89883	conserved hypothet
32	90	11.3	1202	1 S05362	probable DNA-direc
33	89.5	11.2	433	2 A89951	trigger factor [im
34	89.5	11.2	508	2 B81594	hypothetical prote
35	89.5	11.2	508	2 C72074	hypothetical prote
36	89.5	11.2	508	2 E86549	hypothetical prote
37	89.5	11.2	700	2 S67610	probable membrane
38	89.5	11.2	1016	2 T19006	ankyrin related pr
39	89	11.1	528	2 E96795	unknown protein P2
40	89	11.1	1888	2 T39009	hypothetical prote
41	88.5	11.1	301	2 T33088	hypothetical prote
42	88.5	11.1	371	2 A71683	hypothetical prote
43	88.5	11.1	443	2 S66040	serine-type D-Ala-
44	88.5	11.1	762	2 G88436	protein T04A8.13 [
45	88.5	11.1	791	2 T24435	hypothetical prote

ALIGNMENTS

RESULT 1

F95074

serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001

C:Accession: F95074

R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: F95074

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2140 <CUR>

A:Cross-references: KB:AE005672; PIDN:AAK74791.1; PID:g14972117; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0641

Query Match 100.0%; Score 799; DB 2; Length 2140;

Best Local Similarity 100.0%; Pred. No. 7.2e-52;

Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQPDGWEISGFEGKK 60

Db 1953 DTGEVSELKPHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQPDGWEISGFEGKK 2012

Qy 61 DAGVINLSKDTFKPVFKIEEKEENKPTFDVSKKNDPNVHNSQLNESHKEDLQR 120

Db 2013 DAGVINLSKDTFKPVFKIEEKEENKPTFDVSKKNDPNVHNSQLNESHKEDLQR 2072

Qy 121 BEHSQKSDSTKDTATVLDKNNISKSTNNPNK 154

Db 2073 BEHSQKSDSTKDTATVLDKNNISKSTNNPNK 2106

RESULT 2

A97942

metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001

C:Accession: A97942

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

R;Burg, M.A.; Cole, G.J.  
J. Neurobiol. 25, 1-22, 1994  
A>Title: Claustrian, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
A:Reference number: JC5497; MUID:94157526; PMID:7906711  
A:Accession: JC5497  
A:Molecule type: mRNA  
A:Residues: 1-1038 <BUR1>  
A:Cross-references: EMBL:X67778; NID:9406318; PIDN:CAA47988.1; PID:9406319  
A:Accession: PC4334  
A:Molecule type: protein  
A:Residues: 79-83;299-412;485-502 <BUR2>  
A:Experimental source: brain  
C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervous  
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
F:267-270/Region: cell attachment (R-G-D) motif  
F:112,213,490/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 13.6%; Score 108.5; DB 2; Length 1038;  
Best Local Similarity 28.6%; Pred. No. 1.8;  
Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;

Qy 21 KEMSTIVSEDFILPVYGELEKGYQFDGWEISGPEGKDGAVINLSKDTFIKPVFK- 79  
Db 595 KPETKTIIVAKDV-----TTKEQLGKSETSEKQASEKQDVKPKVTKESVKVEVKA 646

Qy 80 KIEEKKEENKPTFDVSKKQNDPQVNHSQLNSHRKEDLQRE-----EHQSQSSTKDV 133  
Db 647 KPEEKDEKEKPKVEVSKEEKPLI---KKEEKPKKEDIKKVEKVEKVEKKEAKKEV 702

RESULT 5  
B71609  
hypoetical protein PFB0680w - malaria parasite (Plasmodium falciparum)  
C:Species: plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: B71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Azavind, L.; Koonin, E.V.,  
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O  
Science 282, 1126-1133, 1998  
A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:199021743; PMID:9804551  
A:Accession: B71609  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown

```

A:Residues: 1-665 <GAR>
A:Cross-references: GB:AE001410; GB:AE001362; NID:g3845245; PIDN:AACT1925.1; PID:g3845242
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0680w

      Query Match      13.1%; Score 105; DB 2; Length 665;
      Best Local Similarity 24.4%; Pred. No. 2;
      Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;

Qy 1 DTGEVSELKPHRVT-VTIQNGKEMSTTIVSEEDFILPVYKGELEKGYQFDGWEI--SGFE 57
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 128 EKNKINKGDLHRQNELNLSQK-----NEQDI-----NKNEKGKQ-----DISNSNAE 170

Qy 58 GKGDAGYVNLKSDTFKPVFKIEBKKE-----EENKPTFD-----VSKCKDNQP 103
      ||||| : : : : : : : : : : : : : : : : : : : : : :
Db 171 NKQ-----VKGEVLELEKKKEEKISDDHKVEENKKSDDHKVEENKKSDDHK 218

Qy 104 VNHSQLNESHKREDLQR-EEHSQKSDSTKDVTVATVLDKNNISSKSTTNNPK 154

```

RESULT 6  
Tl8283  
hypothetical protein G5 - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999



C:Accession: S45817  
R:Favallo, T.  
A:Submitted to the EMBL Data Library, June 1994  
A:Description: The sequence of *S. cerevisiae* cosmid 9205.  
A:Reference number: S46795  
A:Accession: S46817  
A:Molecule type: DNA  
A:Residues: 1-1345 <FAV>  
A:Cross-references: EMBL:U10556; NID:g500825; PIDN:AA68895.1; PID:g500838; MIPS:YHR080C  
C:Genetics:  
A:Cross-references: SGB:S0001122  
A:Map position: 8R  
C:Keywords: transmembrane protein

Query Match 12.2%; Score 97.5; DB 2; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 16;  
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;  
QY 39 KGELEKGYQFDGWEISGFGKK-DAGYVNLKDTFIKVPFKKIEKKEENKPTFDVSK 97  
Db 1109 KGAIKRG-----SVGQKVSVDYMLSELRDII-----SRAKSKPKVKVMK 1149  
QY 98 KKDNPQVNSQLNESHKEDLQREHSOKSDSTKDVATVLD--KNTSSKSTTN--NP 152  
Db 1150 SHDKRPFPKVE-----QKSESRSDDNKDILTHILDFVQNFSSSEIFMKNLLSP 1201  
QY 153 NK 154  
Db 1202 QK 1203

RESULT 11  
T20410  
hypothetical protein E02A10.2 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T20410  
R:Thomas, K.  
A:Reference number: Z19271  
A:Submitted to the EMBL Data Library, October 1996  
A:Accession: T20410  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-385 <WIL>  
A:Cross-references: EMBL:Z81053; PIDN:CAB02877.1; GSPDB:GN00023; CESP:E02A10.2  
A:Experimental source: clone E02A10  
C:Genetics:  
A:Gene: CESP:E02A10.2  
A:Map position: 5  
A:Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

Query Match 12.1%; Score 97; DB 2; Length 385;  
Best Local Similarity 29.6%; Pred. No. 4.3;  
Matches 42; Conservative 24; Mismatches 48; Indels 28; Gaps 7;  
QY 1 DTGEVSELKPHRVTVTIQ--NGKEMSTTVSEDFILPVYKGELEKGYQFDGWEISGREG 58  
Db 256 ETDDFCSLQKENVCHVTILRNHKEVAEK--NEEDKKEEPKKEEKEVEKKE--EDE 310  
QY 59 KQDAGVNLKDTFIKVPFKKIEKKEENKPTFDVSKKKNPQVNSQLNESHKEDL 118  
Db 311 KKDE-----EP--KKEEKKKEEKE--EVEKKER-----BEKKDEEPKKGEE 349  
QY 119 QREHSOKSDSTKDVATVLDK 140  
Db 350 KKEEKEKDEVEEKSEKVEEK 371

RESULT 12  
T14188  
hypothetical protein T28D5.30 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cross)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 28-Jul-2000

C:Accession: T14188  
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro  
submitted to the Protein Sequence Database, August 1999  
A:Reference number: Z17931  
A:Accession: T14188  
A:Molecule type: DNA  
A:Residues: 1-988 <BEV>  
A:Cross-references: EMBL:AL109819  
A:Experimental source: cultivar Columbia; BAC clone T28D5  
C:Genetics:  
A:Gene: ATSP:T28D5.30  
A:Map position: 4  
C:Superfamily: *Arabidopsis thaliana* hypothetical protein T21C14.40

Query Match 12.1%; Score 97; DB 2; Length 988;  
Best Local Similarity 22.1%; Pred. No. 12;  
Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;  
QY 19 NGKEMSTTVSEDFILPVYKGELEKGY-----QPDGWEISGFEKGDAGYVNLKDTF 73  
Db 383 NGRQNSNVQSSVDIELSYTDKVPKSGVGLNVSEIDVELVEDDVRSGAGLSPNVQDN- 441  
QY 74 IKPVFKTIEKKE-----EENKPTFDVSKKKNPQVNSQLNESHKEDLQREE 122  
Db 442 VEPVGDDVRSRSGDMPNPSPANNVREGPATFDIMESEDNPGRDNVAPMEDHIRSEVQLSP 501  
QY 123 HSQKSDSTKDVATVLDKNNISKSTTNPNK 154  
Db 502 HVL---GAKDVTVDSDPTDKVGVNDVTASDPT 532

RESULT 13  
D86432  
hypothetical protein T518.14 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cross)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C:Accession: D86432  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: D86432  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-540 <STO>  
A:Cross-references: GB:AE005172; NID:94587525; PIDN:AAD25756.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 12.1%; Score 96.5; DB 2; Length 540;  
Best Local Similarity 24.8%; Pred. No. 6.9;  
Matches 36; Conservative 25; Mismatches 59; Indels 25; Gaps 5;  
QY 4 EVSELKPHRVTVTIQNGKEMSTTVSEDFILPVYKGELEKGYQPDGWEISGFEKGDAG 63  
Db 39 EEDESKP-----EGVEKSGASFKESDFFADLKESEKK-----ALSCLKSCLBEA 82  
QY 64 YVNLKSDTFFIKVPFKK---TEEKKEENKPTFDVSKKKNPQVNSQLNESHKEDLQ 120  
Db 83 IVDN---TLTKTKKSSPMKKEKKEVEKKEE--EAEKVEKKESEAVVT 136  
QY 121 EHSOKSDSTKDVATVLDKNNIS 145  
Db 137 EAPKAEVTEAVTVEIIPKEVTT 161



Db 2281 LEEEE---KSDKKRD-----DKKNDNTRKNNLDNK 2308

RESULT 14

T37189

hypothetical protein C02H7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000

C:Accession: T37189

R:Leimbach, D.; Minx, M.

A:Submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid C02H7.

A:Reference number: Z20523

A:Accession: T37189

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-535 <LEI>

A:Cross-references: EMBL:U49945; PIDN:AAC47924.1; GSPDB:GNO00029; CESP:C02H7.1

A:Experimental source: strain Bristol N2; clone C02H7

C:Genetics:

A:Gene: CESP:C02H7.1

A:Map position: X

A:Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 12.0%; Score 96; DB 2; Length 535;

Best Local Similarity 21.3%; Pred. No. 7.4;

Matches 35; Conservative 28; Mismatches 61; Indels 40; Gaps 5;

Qy 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKK 60

Db 77 DQGLKVNKAAILI-----SGDAET-----NKMQLMGTNATSFNSRN 116

Qy 61 DAGYVINLSKDTFIKVPFKIEEKEENKPTFDVSKKDNQPNVHSQLNESHKREDLQR 120

Db 117 GTG-----BEKKKKVKEDKKGDEEKST---TKRSSKTHEEKEKSKKSAEE 166

Qy 121 EHSQKSDSTK-----DVTATVLDKNNISSKSTNNPNK 154

Db 167 KEKKKSSSSSRHKSRSRSEKSKSEKSKKKEKSTTDEKPK 210

RESULT 15

G71609

hypothetical protein PFB0650w - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C:Accession: G71609

R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
; Partea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: G71609

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2500 <GAR>

A:Cross-references: GB:AE001408; GB:AE001362; NID:g3845238; PIDN:AAC71919.1; PID:g384524

A:Experimental source: clone 3D7

C:Genetics:

A:Gene: PFB0650w

Query Match 12.0%; Score 96; DB 2; Length 2500;

Best Local Similarity 26.1%; Pred. No. 42;

Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;

Qy 9 KPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINL 68

Db 2173 KPYKIT---ENKK-----NEGNILKKYSIENEKKNYDKEQNECILDKDTQCNVNT 2223

Qy 69 -----SKDTFKPVFKIEEKEENKPTFDVSKKDNQPNVHSQLNE-----SHRKED 117

Db 2224 KEKNLNDKSKSPFNKVKLEEEKSKDKRD---DKKNDNTRKNNLDNKSFPSNIVK 2280

Qy 118 LQREHHSQKSDSKDVTATVLDKNNISSKSTNNPNK 154

Db 119 LQREHHSQKSDSKDVTATVLDKNNISSKSTNNPNK 154

Db 2281 LEEEE---KSDKKRD-----DKKNDNTRKNNLDNK 2308

RESULT 16

T47835

hypothetical protein T209.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T47835

R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Stert, W.; Holland, R.; Weichselgartner, M.;

A:Submitted to the Protein Sequence Database, February 2000

A:Reference number: Z24475

A:Accession: T47835

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-644 <NYA>

A:Cross-references: EMBL:AL138658

A:Experimental source: cultivar Columbia; BAC clone T209

C:Genetics:

A:Map position: 3

A:Introns: 158/2; 329/3

A:Note: T209.90

Query Match 12.0%; Score 95.5; DB 2; Length 644;

Best Local Similarity 23.4%; Pred. No. 10;

Matches 34; Conservative 22; Mismatches 58; Indels 31; Gaps 4;

Qy 18 QNGKMSSTIVSEE---DFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTF 73

Db 519 ENSKTEKTVADKKKSVADFLKRIKNSPQK-----ETTSKKNQKNDGNV----- 565

Qy 74 IKPVFKIEEKEENKPTFDVSKKDNQPNVHSQLNESHKREDLQREHHSQKSDS----- 129

Db 566 -----KKENDHQKSDGNVKKENSKVYKPRELRSSTGKKKVEVNNNSKSSSKRKQ 615

Qy 130 TKQVTVLDKNNISSKSTNNPNK 154

Db 616 TKETAEVATGKRGESGKDKQPRK 640

RESULT 17

I37271

cylicin II - human

C:Species: Homo sapiens (man)

C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 21-Jul-2000

C:Accession: I37271; S52774

R:Hess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.

A:Exp. Cell Res. 218, 174-182, 1995

A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: the

A:Reference number: I37271; MUID:95255491; PMID:7737358

A:Accession: I37271

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-348 <HES>

A:Cross-references: EMBL:Z46788; NID:g758586; PIDN:CAA86752.1; PID:g758587

Query Match 11.9%; Score 95; DB 2; Length 348;

Best Local Similarity 30.9%; Pred. No. 5.4;

Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

Qy 39 KGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKIEEKEENKPTF---DV 95

Db 205 ESEGEKG-----GTEKSKKGGKDS-----KKGKDSALIELQAVKADEKDEDGKDANKGDE 256

Qy 96 SK---KKDNQPNVHSQLN-----ESHKREDLQREHHSQKSDSKD---VTATVLDKNNI 143

Db 257 SKDAKDAKEIKGGKDKKKPSSTDSKDDVKE---SKDATKDAKKVAKDTEKESA 313

Qy 144 SSK 146

Db 314 DSK 316

RESULT 18  
 A09570  
 lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)  
 C:Species: Mycoplasma pulmonis  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001  
 C:Accession: A90570  
 R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;  
 Nucleic Acids Res. 29, 2145-2153, 2001  
 A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm  
 A:Reference number: A99512; MUID:21267165; PMID:11353084  
 A:Accession: A90570  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-622 <KUR>  
 A:Cross-references: GB:AL445566; PID:g14089879; PIDN:CAC13638.1; GSPDB:GN00153  
 A:Experimental source: strain UAB CTIP  
 C:Genetics:  
 A:Gene: MYPU 4650  
 A:Genetic code: SGC3

Query Match	11.9%	Score 95	DB 2	Length 622
Best Local Similarity	25.4%	Pred. No. 10		
Matches	34	Conservative	26	Mismatches 50
				Indels 24
				Gaps 5
Qy	28	VSEEDFTLPVYKGELEKGYQDGEWISGFEFGKDGAGVYINLSKDTFKIPVKLEEEK	87	
Db	85	LSEDNIIIFSLNIENNAKFDLDEF-----VSKDEKFIKFOEINFSQTE	128	
Qy	88	ENKPTFDVSKKCO-----NPWNHSQLNES--HRKEDIQREHSOKSDSTKDVATVLDKN	141	
Db	129	Q-KITDNISKEDEKNKPKDNENNNNSDDQKNDELQKNNSDKLNDNVODEKANKNSN	187	
Qy	142	-NISSSKSTTNPNK	154	
Db	188	SDNSKEKNDENTNK	201	

RESULT 19  
T10466  
DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - malaria parasite (Plasmodium falciparum)  
C/Species: Plasmodium falciparum  
C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000  
C/Accession: T10466  
R;Cheesman, S.J.  
submitted to the EMBL Data Library, September 1995  
A/Reference number: Z17031  
A/Accession: T10466  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-1397 <CHE>  
A/Cross-references: EMBL:X79345; NID:g994807  
C/Genetics:  
A/Gene: TopoII  
A/Map position: 14  
C/Superfamily: eukaryotic type II DNA topoisomerase; phase T4 DNA topoisomerase (ATP-hydrolyzing)  
C/Keywords: ATP; DNA binding; isomerase; nucleus

```

Query Match      11.9%; Score 95; DB 2; Length 1397;
Best Local Similarity 23.8%; Pred. No. 26;
Matches 36; Conservative 36; Mismatches 53; Indels 26; Gaps 6;

Qy      19  NGKEMSTVIYSEDFIL--PVYKGELEKGYQFDGWEISGPEGKDGAGVYVNLKSDFIKP 76
Db      1132 NEEIIAGITVKQDYLLSMPIFSUTLEK---VEDLITQLKEKRELEILRNITVETMWLK 1188

Qy      77  VFVKIEB-----KKEEENKPFVDYSKKQDNQVNHQSOLNESHRKEDLQRBEHSQ 125
Db      1189 DIEKVEEAIEFQRNVELSNEESNK--FKVARKQ-----QPSMMKKKKKKLLSSDESE 1241

Qy      126  ---KSOSTKDVATVLDKNNISSKSTNNPN 153

Db      1242 GGTQSSDSSEFLVNTLNIKQVTKNKTTSNN 1272

```

RESULT 20  
T18427  
hypothetical protein C0335c - malaria parasite (Plasmodium falciparum)  
C/Species: Plasmodium falciparum  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C/Accession: T18427  
R/Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, August 1997  
A/Reference number: Z18935  
A/Accession: T18427  
A/Status: preliminary; translated from GB/EMBL/DBDJ  
A/Molecule type: DNA  
A/Residues: 1-3724 <LAW>  
A/Cross-references: EMBL:Z98547; NID:e1325376; PID:e1325379; PIDN:CAB11104.1  
C/Genetics:  
A/Introns: 307/1; 1545/2  
A/Note: C0335c

```

Query Match      11.8%; Score 94.5; DB 2; Length 3724;
Best Local Similarity 22.1%; Pred. No. 86;
Matches 32; Conservative 30; Mismatches 50; Indels 33; Gaps 6;

Qy      11 HRTVTITQNGKSNSTIVSEDFILPYKGELEKGYQFDGWEISGPEGKKGAGVWNLK 70
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1046 HKIEQDLDIHSQTWICDNN-IEQNEENSCKKGVRISTDW---ENKND----- 1092
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      71 DTETKPVFKIEBKBEENKPTDVSCKDKONPWNH-SQLNESHKEDLQREHSKSDS 129
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1093 -----MENKNDMEKN--DMEKNDIEKNDMEKNDMEKNDMEKNDMEKND 1140
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      130 TKDVTATVLDKNISKSTNNPNK 154
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1141 ME-----NENWENKSDIENENK 1158
      . : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 21
T28771
hypothetical protein E03H12.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T28771
R:Nelson, J.; Wohlmann, P.; Sansone, J.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of C. elegans cosmid E03H12.
A:Reference number: Z20520
A:Accession: T28771
A:Status: preliminary; translated from CB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-210 <N80>
A:Cross-references: EMBL:AF000299; PIDN:AAAC47980.1; GSPDB:GN00022; CESP:E03H12.
A:Experimental source: strain Bristol N2; clone E03H12
C:Genetics:
A:Gene: CESP:E03H12.5
A:Map position: 4
A:Introns: 30/2: 201/3

```

```
Query Match      11.8%; Score 94; DB 2; Length 210;  
Best Local Similarity 28.6%; Pred.No. 3.6;  
Matches 28; Conservative 18; Mismatches 52; Indels 0; Gaps 0;
```

Qy	57	EKKDAGYVNLKDTFLPKVFVKI	BKKKEENKNPTFDSKKDNQVNHSQNLSHRKE	116
		:   :	:   :   :   :   :   :   :	
Dd	69	EGEKDGDKSKSEKKGDKKEEEKDEBKDGDKEDDKDEKDDEKDAEKKOE	KD	128
		:   :	:   :   :   :   :   :   :	
Qy	117	DLOREHESQKSDTKDVATATVL	DKNNISKSSTTNNPNK	154
		:   :	:   :   :   :   :   :   :	
Dd	129	EKKDDKDEKDEKDEKDEKSKS	KSSKSSKSSKSSKLKSLK	166
		:   :	:   :   :   :   :   :   :	

RESULT 22  
T05612  
hypothenical protein Fgpl6.270 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
C:Accession: T05612  
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohelsel, J.; Mewes, H.W.; Mayer, K.F.  
submitted to the Protein Sequence Database, February 1999  
A:Reference number: Z15419  
A:Accession: T05612  
A:Molecule type: DNA  
A:Cross-references: EMBL:AL035394  
A:Experimental source: cultivar Columbia; BAC clone F9D16  
C:Genetics:  
A:Map position: 4  
A:Introns: 110/3; 247/2; 282/3; 304/3; 361/3; 390/2; 418/3  
A:Note: F9D16.270

Query Match 11.7%; Score 93.5; DB 2; Length 456;  
Best Local Similarity 25.0%; Pred. No. 9.5;  
Matches 39; Conservative 28; Mismatches 58; Indels 31; Gaps 8;

QY 4 EVSELPK-HRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDA 62  
DB 249 EKDPKPKHPSAFIVYANERAAALREENKSVFAK-----ITGEENKNSDKKA 300  
QY 63 GY--VINLSKDTFIKPVFKIEKKEENKPTFDVSKKNDPNQVNHSQLNESHKEDLQR 120  
DB 301 PYEKVAKKNKETYLQ-AMEEYKRTKEE-----ALSQKE-----EBELLKLHQEALQM 349  
QY 121 EHSQKSDSTKDVTVATVLDKNISSKSTNN--PNK 154  
DB 350 LKKEKTDN-----LIKKEKATKKKKNVNDPNK 378

RESULT 23  
B72291  
hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: B72291  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571  
A:Accession: B72291  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-219 <ARN>  
A:Cross-references: GB:AE001771; GB:AE000512; NID:g4981678; PIDN:AAD36218.1; PID:g498168  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM1142

Query Match 11.6%; Score 93; DB 2; Length 219;  
Best Local Similarity 26.2%; Pred. No. 4.5;  
Matches 42; Conservative 28; Mismatches 50; Indels 40; Gaps 9;

QY 8 LKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQ--FDGWEISG-----PEGK 59  
DB 64 LIPNVFVELYATDEKTLFAKE-----VLGEESVSRYDLFAGFGVGRGTPPTFFFGK 116  
QY 60 KDAGVIN-LKDTPIKPVFKIEKKEENKPTFDVSKKNDPNQVNHSQLNESHKED- 117  
DB 117 EGLGYPGVKDNFIK-ILKYVAQELKED----FQTYLKKKDDPPVGPFLIETP-KEDA 170  
QY 118 ---LQREHSQKSDS-----TKDVTVATVLDK 140  
DB 171 DFVLEKDNVAVKDVTPNEVRDRRIYVTSDPVAKTLQEK 210

hypothetical protein C17F3.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T32879  
R:Gattung, S.; Scheet, P.  
submitted to the EMBL Data Library, January 1998  
A:Description: The sequence of C. elegans cosmid C17F3.  
A:Reference number: Z21240  
A:Accession: T32879  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-253 <GAT>  
A:Cross-references: EMBL:AF043692; PIDN:AAB97531.1; GSPDB:GN00019; CESP:C17F3.3  
A:Experimental source: strain Bristol N2; clone C17F3  
C:Genetics:  
A:Gene: CESP:C17F3.3  
A:Map position: 1  
A:Introns: 41/1

Query Match 11.4%; Score 91; DB 2; Length 253;  
Best Local Similarity 32.6%; Pred. No. 7.6;  
Matches 31; Conservative 7; Mismatches 33; Indels 24; Gaps 3;

QY 76 PVFKIEKKE---EENKPTFDVSKK-----KDNQVNHSQLNESHK----- 114  
DB 61 PVAPKVEKKEEKEEKADDEKKTEEKDDKSKKTEEKDKISVKTKTQETSERKDKK 120  
QY 115 ---KEDLQREHSQKSDSTKDVTVATVLDKNISSK 146  
DB 121 DERKEDEKKESEKSKDEEKKDKDEKDEKK 155

RESULT 25  
G96542  
hypothetical protein F17J6.14 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: G96542  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G96542  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-629 <STO>  
A:Cross-references: GB:AE005173; NID:g11054631; PIDN:AAG27876.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F17J6.14  
A:Map position: 1

Query Match 11.3%; Score 90.5; DB 2; Length 629;  
Best Local Similarity 19.5%; Pred. No. 23;  
Matches 33; Conservative 37; Mismatches 66; Indels 33; Gaps 5;

QY 10 PHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDGAVINLS 69  
DB 423 PHLETAKPTKDSMEQTEAEADVAMNPIVEKAMSEVMEAGAAINPIVEADG-----A 476  
QY 70 KDTPIKPVFKIEKKE-----EENKPTFDVSKKNDPNQ--VNHSQLNESHKKE 116  
DB 477 MNPIVEKAMSOIVAEADAAINQAVDFQTAQTGNDDAESDDFSEPVSHS---ETLNPP 533  
QY 117 DLQREHSQKSDSTKDVTVATVLDKNN-----ISSKSTTNNPNK 154  
DB 534 ELEKKEVWRKDATERSVSADCCQDKNSKIABESSLQEEISASQTSNPTQ 582

## RESULT 26

T28391  
ORF MSV230 hypothetical protein - Melanoplus sanguinipes entomopoxvirus  
C:Species: Melanoplus sanguinipes entomopoxvirus  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T28391  
R:Afonso, C.L.; Tulman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: Z20484; MUID:99102612; PMID:9847359  
A:Accession: T28391  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-670 <AFO>  
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97713.1; PID:g4049753  
C:Genetics:  
A:Note: MSV230

Query Match 11.3%; Score 90.5; DB 2; Length 670;  
Best Local Similarity 22.5%; Pred. No. 25;  
Matches 36; Conservative 34; Mismatches 69; Indels 21; Gaps 5;  
  
QY 15 VTIQKEMSSIT- - - - - VSEEDFILPVYKGELEKGY- - - - - QFDGWEISGF 56  
DB 33 VNFEEKQIITLLKFNFDKTEMCGVSEKVFQLINNKASAEKYSVDVDSIDESQNSDS 92  
QY 57 EGKDGAGYVILNSKDTFIKPVFKIE- - - - - EKKEENKPTFDVSKKKNPQVNHSQL- - - - - NESHR 114  
DB 93 DSDSDSGVNIQNSDSKVNIKLENESQNSDSKVNIQNSDS- - - - - KVINIKLENESQ 151  
QY 115 KEDLQREHSQKSDSTKYVTATVLDKNNISSKSTTNPNK 154  
DB 152 SDSKVNIQNSDSKVNIQNSDSKVNIQNSDSK 191

## RESULT 27

T27136  
hypothetical protein Y53C12B.3a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T27136  
R:Kershaw, J.; Lennard, N.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z20316  
A:Accession: T27136  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-867 <WIL>  
A:Cross-references: EMBL:Z99278; PIDN:CAB16493.1; GSPDB:GN000020; CESP:Y53C12B.3a  
A:Experimental source: clone Y53C12B  
C:Genetics:  
A:Gene: CESP:Y53C12B.3a  
A:Map position: 2  
A:Introns: 100/3; 177/3; 218/1; 423/3; 714/2; 864/3

Query Match 11.3%; Score 90.5; DB 2; Length 867;  
Best Local Similarity 26.4%; Pred. No. 33;  
Matches 42; Conservative 27; Mismatches 57; Indels 33; Gaps 7;  
  
QY 4 EVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELE- - - - - KGYQFDGWEISGFEGK 59  
DB 672 EISEENPK--TDDIQSKDDVTS- - - - - KSELHCYRCHEYQLPAEEVSSHNR 716  
QY 60 KDAGYVILNSKDTFIKPVFKIE- - - - - EKKEENKPTFDVSKKKNPQVNHSQLNESHRK 115  
DB 717 KONG- - - - - DLWCCEHM--KKIKGHCCEATGEGHPLICPKKKEERVAKSR--ESSQK 766  
QY 116 EDLQREHSQKSDSTKYVTATVLDKNNISSKSTTNPNK 154  
DB 767 PIDPQEISDDQDDTVDPDQIVEQDNQSHKSRHNSNR 805

## RESULT 28

T27135  
hypothetical protein Y53C12B.3b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T27135  
R:Kershaw, J.; Lennard, N.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: Z20316  
A:Accession: T27135  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-871 <WIL>  
A:Cross-references: EMBL:Z99278; PIDN:CAB16492.1; GSPDB:GN000020; CESP:Y53C12B.3b  
A:Experimental source: clone Y53C12B  
C:Genetics:  
A:Gene: CESP:Y53C12B.3b  
A:Map position: 2  
A:Introns: 100/3; 177/3; 218/1; 423/3; 717/2; 867/3

Query Match 11.3%; Score 90.5; DB 2; Length 871;  
Best Local Similarity 26.4%; Pred. No. 33;  
Matches 42; Conservative 27; Mismatches 57; Indels 33; Gaps 7;  
  
QY 4 EVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELE- - - - - KGYQFDGWEISGFEGK 59  
DB 675 EISEENPK--TDDIQSKDDVTS- - - - - KSELHCYRCHEYQLPAEEVSSHNR 719  
QY 60 KDAGYVILNSKDTFIKPVFKIE- - - - - EKKEENKPTFDVSKKKNPQVNHSQLNESHRK 115  
DB 720 KONG- - - - - DLWCCEHM--KKIKGHCCEATGEGHPLICPKKKEERVAKSR--ESSQK 769  
QY 116 EDLQREHSQKSDSTKYVTATVLDKNNISSKSTTNPNK 154  
DB 770 PIDPQEISDDQDDTVDPDQIVEQDNQSHKSRHNSNR 808

## RESULT 29

S41552  
probable transcription factor SPT7 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein YBR0739; protein YBR081C  
C:Species: Saccharomyces cerevisiae  
C:Date: 28-Jan-1994 #sequence\_revision 09-Sep-1994 #text\_change 20-Sep-1999  
A:Accession: S41552; S45946; S45948; S40800; S45478; S54985; S59716  
R:Gansheroff, L.; Dollard, C.; Fan, P.; Winston, F.  
submitted to the EMBL Data Library, July 1993  
A:Reference number: S41552  
A:Accession: S41552  
A:Molecule type: DNA  
A:Residues: 1-1332 <GAN>  
A:Cross-references: EMBL:L22537; NID:g349189; PIDN:AAC37424.1; PID:g349190  
R:Steensma, H.Y.; van der Aart, Q.J.M.  
submitted to the Protein Sequence Database, August 1994  
A:Reference number: S45932  
A:Accession: S45946  
A:Molecule type: DNA  
A:Residues: 1-1332 <STE>  
A:Cross-references: EMBL:Z35950; NID:g536341; PIDN:CAAB5026.1; PID:g536342; MIPS:YBR081C  
R:Andre, B.; Cziepluch, C.; Hein, C.; Jauniaux, J.C.; Urrestazu, A.; Viissers, S.  
submitted to the Protein Sequence Database, August 1994  
A:Reference number: S45893  
A:Accession: S45948  
A:Molecule type: DNA  
A:Residues: 1-835 <AND>  
A:Cross-references: EMBL:Z35950; MIPS:YBR081C  
R:Haynes, S.R.; Dollard, C.; Winston, F.; Beck, S.; Trowsdale, J.; Dawid, I.B.  
Nucleic Acids Res. 20, 2603, 1992  
A:Title: The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins  
A:Reference number: S40800; MUID:92285152; PMID:1350857  
A:Accession: S40800  
A>Status: nucleic acid sequence not shown  
A:Molecule type: DNA

A;Accession: 463-523 <HAY>  
A;Cross-references: EMBL:M87651; NID:gl72683; PIDN:AAA35087.1; PID:gl72684  
R;van der Aart, O.J.M.; Barthe, C.; Doignon, F.; Aigle, M.; Crouzet, M.; Steensma, H.Y.  
Yeast 10, 959-964, 1994  
A;Title: Sequence analysis of a 31 kb DNA fragment from the right arm of Saccharomyces cerevisiae  
A;Reference number: S45462; MUID:95076715; PMID:7985423  
A;Accession: S45478  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-624, 'LRGKKRKI', 633-1332 <VAN>  
A;Cross-references: EMBL:X76294  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993  
R;Gansheroff, L.J.; Dollard, C.; Tan, P.; Winston, F.  
Genetics 139, 523-536, 1995  
A;Title: The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein important for cell cycle arrest  
A;Reference number: S54985; MUID:95229044; PMID:7713415  
A;Accession: S54985  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-1332 <GAW>  
A;Cross-references: EMBL:L22537; NID:g349189; PIDN:AAC37424.1; PID:g349190  
R;van der Aart, O.J.M.  
submitted to the EMBL Data Library, August 1995  
A;Accession: S59716  
A;Reference number: S59716  
A;Molecule type: DNA  
A;Residues: 1-1332 <VAM>  
A;Cross-references: EMBL:X76294; NID:g974203; PIDN:CAA53940.1; PID:e264674; PID:g558360  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: SGD:SPT7  
A;Cross-references: SGD:S0000285; MIPS:YBR081c  
A;Map position: 2R  
C;Superfamily: unassigned bromodomain proteins; bromodomain homology  
C;Keywords: nucleus; transcription regulation  
F;466-521/Domain: bromodomain homology <BRO>  
  
Query Match 11.3%; Score 90.5; DB 2; Length 1332;  
Best Local Similarity 24.5%; Pred. No. 54;  
Matches 39; Conservative 26; Mismatches 57; Indels 37; Gaps 8;  
  
Qy 15 VTIONGKMSSTIIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINISKDTFI 74  
Db 549 ITRNRADLEKEI---EDM-----EKDKVYELDEEEVAGSRKG---LNNGAHMLA 594  
  
Qy 75 K---PVFKKIEEKEENKPTFD-----VSKKD-----NPQVNHSQLNESH 113  
Db 595 KENGKVSSEKSSKTVKDEAPTNDKLTSVIPEGEKEKDKTASSTVTVHENVNKNKIKENG 654  
  
Qy 114 RKEDLQR--EEHOKSDSTKDVATVLD--KNNISKSTTN 150  
Db 655 KNEEQDMVESSKTEDSSKADAKKOTEDGLQDKTAEN 693  
  
RESULT 30  
T28676  
rhoptry protein - Plasmodium yoelii (fragment)  
C;Species: Plasmodium yoelii  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 01-Dec-2000  
A;Accession: T28676; A45521  
R;Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.  
Mol. Biochem. Parasitol. 76, 329-332, 1996  
A;Title: Comparison of two members of a multigene family coding for high-molecular mass  
A;Reference number: Z20507; MUID:97077455; PMID:8920022  
A;Accession: T28676  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2401 <SIN>  
A;Cross-references: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:ABA41263.1  
R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.  
Mol. Biochem. Parasitol. 42, 241-246, 1990  
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple clones  
A;Reference number: A45521; MUID:91101660; PMID:2270106

A;Accession: A45521  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 2260-2401 <KEE>  
A;Cross-references: GB:M34281  
  
Query Match 11.3%; Score 90.5; DB 2; Length 2401;  
Best Local Similarity 27.4%; Pred. No. 1e+02;  
Matches 43; Conservative 26; Mismatches 53; Indels 35; Gaps 6;  
  
Qy 16 TIQNGKMSSTI-----VSEEDFILPVYKGELEKGY---QFDGWEISGFEGKDA 62  
Db 924 TSKNHEKSIDRKNSLKIQDFSESYINDI-KKELEKNVLESQNNNTDINQLSKEN 982  
  
Qy 63 GYVINLSKDTFIKPVFKKIEEKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQREE 122  
Db 983 IY--NILKLNKIKKIDKVKETDEIK-----NNKINAELSSEKIITOLKENSILKE 1035  
  
Qy 123 HSQKSDSTK-----VTATVLDKNMISS 145  
Db 1036 CQSIKSTIDNYYVSECIKNITNLKTYIVNEKNNT 1072  
  
RESULT 31  
B89883  
conserved hypothetical protein SA0976 [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
A;Accession: B89883  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguchi, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: B89883  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-645 <KUR>  
A;Cross-references: GB:BA000018; PID:g13700929; PIDN:BAB42225.1; GSPDB:GN00149  
A;Experimental source: strain N315  
C;Genetics:  
A;Gene: SA0976  
  
Query Match 11.3%; Score 90; DB 2; Length 645;  
Best Local Similarity 21.4%; Pred. No. 26;  
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;  
  
Qy 4 EVSELKPHRTV--TIQNGKMSSTIIVSEEDFILPVYKGELE-KGYQF-----DGWEIS 54  
Db 341 KMTDLQDTKYVYVESVENNESMDTFVKH-----PIKTGMLNGKKYVMVMTTNDYWKDF 395  
  
Qy 55 GFEGKK-----DAGYVINL-SKDTFI 74  
Db 396 MVEGQRVTTISKDAKNNTRTIIFPVYEGKTLTYDAIVKHVVKTYDYGQYHVRIVDKFAFT 455  
  
Qy 75 KPVFKKIEEKEENKPTFDV-----SKKDNPNQVNHSQLNESHKEDLQ---REHSQ 125  
Db 456 KANTDKSNKKEQQQDNKAKKATPATPKPTSPVPEKESQKDSQKDDNKQKLPVSEKENDA 515  
  
Qy 126 KSDSTKDVATVLDKNNISKSTTNPNK 154  
Db 516 SSESQDKTPTATPKTKEVESSSIT--PTK 543  
  
RESULT 32  
S05362  
probable DNA-directed DNA polymerase (EC 2.7.7.7) - fungus (Ascobolus immersus) mitochondrion  
C;Species: mitochondrion Ascobolus immersus  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
A;Accession: S05362  
R;Kempken, F.; Meinhardt, F.; Esser, K.  
Mol. Gen. Genet. 218, 523-530, 1989



A:Reference number: A86491; MUID:20330349; PMID:10871362  
A:Accession: E86549  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-508 <STO>  
A:Cross-references: GB:BA000008; NID:98978843; PIDN:BA98679.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: CPJ0473

Query Match 11.2%; Score 89.5; DB 2; Length 508;  
Best Local Similarity 24.5%; Pred. No. 21;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
Qy 74 IKPVFKTEKEENKPTFD-----VSKKDNQVNHSLNESHRK 115  
Db 92 VGVFVKTPQARPEVSSRLPSHVQGLPGLGFRDRIQKRSNPADLGKMKRSYSD 151  
Qy 116 EDLQREHSQKSDTKDVTATVLDKNNISSKSTT 149  
Db 152 GDLDRVGHDSNEDSTEDSRK---EGGEPSSKSS 182

RESULT 37  
S67610  
probable membrane protein YDL074c - Yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein D2483  
C:Species: Saccharomyces cerevisiae  
C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002  
R:Wambutt, R.; Wedler, H.; Wedler, E.; Scharfe, M.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67608  
A:Accession: S67610  
A:Molecule type: DNA  
A:Residues: 1-700 <WAM>  
A:Cross-references: EMBL:Z74122; NID:G1431087; PID:G253213; PID:G1431088; GSPDB:GN00004;  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:BRE1; MIPS:YDL074C  
A:Cross-references: SGD:S0002232  
A:Map position: 4L  
C:Keywords: transmembrane protein  
P:69-85/Domain: transmembrane #status predicted <TMM>

Query Match 11.2%; Score 89.5; DB 2; Length 700;  
Best Local Similarity 23.5%; Pred. No. 31;  
Matches 36; Conservative 29; Mismatches 47; Indels 41; Gaps 7;  
Qy 19 NGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKK----- 60  
Db 430 NDTKSSST-----QDALIKEIQ-DLEKGR-----ELSDLTHKKYSIINHESVISKLTVK 480  
Qy 61 ---DAGYVINL-SKDTF---IKPVFKTEKEENKPTFDVSKKDNQVNHSLNESHR 113  
Db 481 TKADQKYFAAWRSKDSILIEIKLSLSKSNEL-----ILQLKSDRLLOQKIGNLH 533  
Qy 114 RKEDLQREHSQKSDTKDVTATVLDKNNISSK 146  
Db 534 KOLDLSONNERRLIDSSKTETLKIIDLANTSTK 566

RESULT 38  
T19006  
ankyrin related protein C06C3.1 - Caenorhabditis elegans  
N:Contains: myosin-light-chain-phosphatase (EC 3.1.3.53)  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Mar-2000  
C:Accession: T19006; T22086  
R:Berks, M.  
submitted to the EMBL Data Library, August 1994  
A:Reference number: Z19058  
A:Accession: T19006

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1016 <WIL>  
A:Cross-references: EMBL:Z36719; PIDN:CAA85318.1; GSPDB:GN00020; CESP:C06C3.1  
A:Experimental source: clone C06C3  
R:Matthews, P.  
submitted to the EMBL Data Library, January 1995  
A:Reference number: Z19510  
A:Accession: T22086  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1016 <WIL2>  
A:Cross-references: EMBL:Z47809; PIDN:CAA87782.1; GSPDB:GN00020; CESP:C06C3.1  
A:Experimental source: clone F42A8  
C:Genetics:  
A:Gene: CESP:C06C3.1  
A:Map position: 2  
A:Introns: 27/3; 94/3; 279/3; 352/2; 462/2; 523/3; 569/2; 718/3; 766/3; 833/3; 89/  
C:Keywords: phosphoric monoester hydrolase

Query Match 11.2%; Score 89.5; DB 2; Length 1016;  
Best Local Similarity 21.6%; Pred. No. 47;  
Matches 33; Conservative 33; Mismatches 38; Indels 49; Gaps 8;  
Qy 15 VTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSKOTFI 74  
Db 265 ILLENGAELSD-----LTFGADVGLGVADKECIDYVELA-DTV- 302  
Qy 75 KPVPFKTEKEEER---NKPTFDVSKKDNQVNHSLNESHRKEDLQ-REESQK--- 126  
Db 303 ---KVQNKTKSPGSGSQPTSLQEKNHMPSEEHVLTSEKRDQLQHKDQHSNBL 357  
Qy 127 -----SDSTKDVAT-VLDKNNISS 145  
Db 358 HSHFSTASVGSTTSNTNTTTTIVIGENDISA 390

RESULT 39  
E96795  
unknown protein F28016.8 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: E96795  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: E96795  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-528 <STO>  
A:Cross-references: GB:AE005173; NID:G6143888; PIDN:AAF04434.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F28016.8  
A:Map position: 1

Query Match 11.1%; Score 89; DB 2; Length 528;  
Best Local Similarity 24.3%; Pred. No. 24;  
Matches 35; Conservative 22; Mismatches 51; Indels 36; Gaps 6;  
Qy 41 ELEGXGYPDGW-----EISGFEGKKDAG-----YVINLSKDTFTKVPFKTEE 83  
Db 200 ELAYDINFEWYGGAKVRCLCGAVACSGFLGKSGFQASPVIIIEDTY---VWEDGDD 256  
Qy 84 KKEENKPTFDVSKK---DNQVNHSLNESHRKEDLQREEH-----SQKSDTKD- 132



Search completed: February 10, 2004, 10:58:34  
Job time : 12.1197 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 7.15925 Seconds  
(without alignments)  
1011.574 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799

Sequence: 1 DTGEVSELKPHRVTTIQNG.....ATVLDKNISSKSTNNPNK 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97.5	12.2	1345	1 YH00_YEAST	P38800 saccharomyc
2	95	11.9	348	1 CYL2_HUMAN	Q14093 homo sapien
3	95	11.9	1398	1 TOP2_PLAFLK	P41001 plasmodium
4	92.5	11.6	2468	1 MAPB_HUMAN	P46821 homo sapien
5	90.5	11.3	1332	1 SP7_YEAST	P35177 saccharomyc
6	90	11.3	1202	1 DPOM_ASCIM	P22374 ascobolus i
7	89.5	11.2	433	1 TIG_STAAM	Q99316 staphylococ
8	89	11.1	1888	1 YDT2_SCHPO	O14207 schizosacch
9	88.5	11.1	443	1 DACA_BACSU	P08750 bacillus su
10	87.5	11.0	1702	1 IGA2_HAEIN	P45384 haemophilus
11	87	10.9	893	1 GYRA_STAEP	P54112 staphylococ
12	87	10.9	2464	1 MAPB_MOUSE	P14873 mus musculu
13	86.5	10.8	1006	1 RAT1_YEAST	Q02792 saccharomyc
14	86.5	10.8	1875	1 MLP1_YEAST	Q02455 saccharomyc
15	86	10.8	443	1 GLNA_PYRKO	O08467 pyrococcus
16	86	10.8	5596	1 MDN1_HUMAN	Q9nu22 homo sapien
17	85.5	10.7	914	1 PBPA_BACSU	P39793 bacillus su
18	85.5	10.7	1007	1 RF21_YEAST	P39083 saccharomyc
19	85	10.6	720	1 IGAI_STAEP	Q8cst4 staphylococ
20	85	10.6	6632	1 UN89_CAEEL	O01761 caenorhabdi
21	84.5	10.6	1694	1 IGA2_HAEIN	P44969 haemophilus
22	84	10.5	778	1 YFK8_YEAST	P43610 saccharomyc
23	83.5	10.5	479	1 U2R1_HUMAN	Q15695 homo sapien
24	83.5	10.5	943	1 ARS2_DROME	Q9v9k7 drosophila
25	83	10.4	258	1 IF31_HUMAN	O75822 homo sapien
26	83	10.4	270	1 TONB_HAEIN	P42872 haemophilus
27	83	10.4	688	1 LIP_STAEP	Q02510 staphylococ
28	83	10.4	715	1 ADSV_HUMAN	Q9y9u3 homo sapien
29	83	10.4	847	1 DNLI_CAEEL	Q27474 caenorhabdi
30	82.5	10.3	439	1 GLNA_PYRAB	Q9uy99 pyrococcus
31	82.5	10.3	649	1 HS70_PYRSA	P37899 pyrenomonas
32	82.5	10.3	2459	1 MAPB_RAT	P15205 rattus norv
33	82	10.3	279	1 YMB4_YEAST	P49957 saccharomyc

34	82	10.3	427	1 YK5_YEAST	P36046 saccharomyc
35	82	10.3	439	1 GLNA_PYRPU	Q05907 pyrococcus
36	82	10.3	439	1 GLNA_PYRMO	P36687 pyrococcus
37	82	10.3	943	1 CENC_HUMAN	Q03188 homo sapien
38	82	10.3	949	1 IF2_HELPJ	Q9zm46 helicobacte
39	82	10.3	1251	1 RBP2_PLAVB	Q00799 plasmodium
40	81.5	10.2	425	1 CRT1_ARATH	O04151 arabidopsis
41	81.5	10.2	572	1 LMD1_HUMAN	P29536 homo sapien
42	81.5	10.2	657	1 Y040_MYCPN	P75062 mycoplasma
43	81.5	10.2	676	1 RLM1_YEAST	Q12224 saccharomyc
44	81.5	10.2	752	1 DRS1_YEAST	P32892 saccharomyc
45	81	10.1	242	1 Y376_UREPA	Q3pqb5 ureaplasma

ALIGNMENTS

RESULT 1

ID	YH00_YEAST	STANDARD;	PRT;	1345 AA.
AC	P38800;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	Hypothetical 149.7 kDa protein in IREL-KSP1 intergenic region.			
GN	YHR080C.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=S288c / AB972;			
RX	MEDLINE=94378003; PubMed=8091229;			
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,			
RA	Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,			
RA	Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,			
RA	Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,			
RA	Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,			
RA	Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,			
RA	Vaudin M.;			
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome			
RT	VIII."			
RL	Science 265:2077-2082(1994).			
CC	-I- SIMILARITY: TO YEAST YFL042C.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL;	U10556;	AAB68895.1;	
DR	PIR;	S46817;	S46817
DR	SGD;	S0001122;	YHR080C.
DR	InterPro;	IPR004182;	GRAM_dom.
DR	Pfam;	PF02893;	GRAM; 1.
DR	SMART;	SM00568;	GRAM; 1.
KW	Hydrophobic protein; Transmembrane.		
FT	TRANSMEM 1198 1218	POTENTIAL.	
SQ	SEQUENCE 1345 AA;	149679 MW;	2FDDB9A4686564C2 CRC64;

Query Match	12.2%;	Score 97.5;	DB 1;	Length 1345;
Best Local Similarity	27.0%;	Pred. No. 7.9;		
Matches	33;	Conservative	18;	Mismatches 38;
				Indels 33;
				Gaps 6;
Qy	39	KGELEKGVQPCWISGEGKK-DAGYVINLSKDTFKVPFKKIEEKEEENKPTFDYSK	97	
Db	1109	KGAIEKG-----SVGQKVSVDYMLSELDII-----	-----	-----
Qy	98	KKDNPQVNHSQLNESHKREDIQREHSQKSDTKDVTATVLD--	-----	-----
				KNNISSKSTTN---NP 152

```
Db 1150 SHDKHPFISKVE-----QKSSSRKDDNKOLLTHILDFVQNNFSSEIFMKNLLSP 1201
QY 153 NK 154
Db 1202 QK 1203

RESULT 2
CYL2_HUMAN
ID CYL2_HUMAN STANDARD; PRT; 348 AA.
AC Q14093;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cyclicin II (Multiple-band polypeptide II).
GN CYL2 OR CYL2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=95355491; PubMed=7737358;
RA Hess H., Heid H., Zimbelmann R., Franke W.W.;
RT "The protein complexity of the cytoskeleton of bovine and human sperm
RT heads: the identification and characterization of cyclicin II.";
RL Exp. Cell Res. 218:174-182(1995).
CC -!- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY
CC BE INVOLVED IN SPERMATID DIFFERENTIATION.
CC -!- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.
CC -!- TISSUE SPECIFICITY: Testis.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z46788; CAA86752.1; -.
CC PIR; I37271; I37271.
CC Genew; HGNC:2583; CYL2.
CC MIM; 604035; -.
CC GO: GO:0005200; F:structural constituent of cytoskeleton; TAS.
KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
FT DOMAIN 25 347 31 X 3 AA REPEATS OF K-K-X.
FT REPEAT 157 240 3 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 157 184 1.
FT REPEAT 185 212 2.
FT REPEAT 213 240 3.
SQ SEQUENCE 348 AA; 39079 MW; D8676599C1809E7 CRC64;

Query Match 11.9%; Score 95; DB 1; Length 348;
Best Local Similarity 30.9%; Pred. No. 2.8;
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 39 KGELEKGYQPDGWEISGFEGKDKAGVINLSKDTFTFKPVFKIEEKEENKPTP----DV 95
Db 205 ESRGEGK----GTEKDSKKGKDS-----KKGKDSALQVADKEDGDKDANKGDE 256
QY 96 SK--KKDNPOVNSQLN-----ESHKEDLQREHSOKSDSTKD---VTATVLDDKNNI 143
Db 257 SKDAKKDAKEIKGKKKKPSSTDSKDDVAKK---SKKATKDAKKVAKDTKEKSA 313
QY 144 SSK 146
Db 314 DSK 316

RESULT 3
```

```
TOP2_PLAFK
ID TOP2_PLAFK STANDARD; PRT; 1398 AA.
AC P41001;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94316496; PubMed=8041616;
RA Cheesman S., McAleese S., Goman M., Johnson D., Horrocks P.,
RA Ridley R.G., Kilbey B.J.;
RA "The gene encoding topoisomerase II from Plasmodium falciparum.";
RL Nucleic Acids Res. 22:2547-2551(1994).
CC -!- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMS
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -!- SIMILARITY: Belongs to the type II topoisomerase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X79345; -. NOT ANNOTATED CDS.
CC HSSP; P06786; 1BGM.
CC InterPro; IPR003594; ATPbind ATPase.
CC InterPro; IPR003957; CBFA_NFYB_topis.
CC InterPro; IPR001241; DNA_topoisoi.
CC InterPro; IPR002205; DNA_topoisoiV.
CC Pfam; PF00204; DNA_gyraseB; 1.
CC Pfam; PF02518; DNA_topoisoiV; 1.
CC Pfam; PF02518; HATPase_c; 1.
CC PRINTS; PR00615; CCAATSUBUNTA.
CC PRINTS; PR00418; TP12FAMILY.
CC ProDom; PD000742; DNA_topoisoiV; 1.
CC SMART; SM00387; HATPase_c; 1.
CC SMART; SM00433; TOP2c; 1.
CC SMART; SM00434; TOP4c; 1.
CC PROSITE; PS00177; TOPOISOMERASE II; 1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
FT NP_BIND 144 149 ATP (POTENTIAL).
FT ACT_SITE 830 830 DNA CLEAVAGE (BY SIMILARITY).
FT DOMAIN 271 281 POLY-ASN.
FT DOMAIN 308 316 POLY-ASN.
FT DOMAIN 1089 1093 POLY-LYS.
FT DOMAIN 1227 1234 POLY-LYS.
SQ SEQUENCE 1398 AA; 161029 MW; BAAD7BEE88FE5BE9 CRC64;

Query Match 11.9%; Score 95; DB 1; Length 1398;
Best Local Similarity 23.8%; Pred. No. 13;
Matches 36; Conservative 36; Mismatches 35; Indels 26; Gaps 6;

QY 19 NGKMSSTIVSEEDFIL--PVYKGELEKGYQPDGWEISGFEGKDKAGVINLSKDTFTFKP 76
Db 1133 NEEIAGITVKDYDYLMSPIFSLTLEK---VEDLLTQKERELEILRNITVETMLK 1189
QY 77 VFKTIEE-----KKEENKPTFDVSKKNPQVNSQLNESHKEDLQREHSQ 125
```

Db 1190 DIEKVEAIEPQNVNELSREESNK--FKVARKQ-----GPSMMKKKKKKKKLSSDEESE 1242

Qy 126 ---KSDTKQVATVLDKNNISSTNNPN 153

Db 1243 GGDTSDSSEFLVNLNKKNTNKKTTTSSNN 1273

RESULT 4

MAP1 HUMAN

AC P46821; STANDARD; PRT; 2468 AA.

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain LC1].

GN MAP1B.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal brain;

RX MEDLINE=95104835; PubMed=7806212;

RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;

RT "Cloning of human microtubule-associated protein 1B and the identification of a related gene on chromosome 15.";

RL Genomics 22:273-280(1994).

CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN. PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN STABILIZING MICROTUBULES.

CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE WITH MAP1A AND MAP1B PROTEINS.

CC -!- DOMAIN: Has a highly basic region with many copies of the sequence KKEE and KKEI/V repeated but not at fixed intervals, which is responsible for the binding of MAP1B to microtubules.

CC -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated from MAP1B by proteolytic processing. It is free to associate with both MAP1A and MAP1B. It interacts with the amino-terminal region of MAP1B (By similarity).

CC -!- SIMILARITY: TO MAP1A.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L06237; AAA18904.1; .

DR Genbank; HGNC:6836; MAP1B.

DR MIM; 157129; .

DR GO; GO:0005875; C: microtubule associated complex; TAS.

DR InterPro; IPR000102; MAP1B neuraxin.

DR Pfam; PF00414; MAP1B neuraxin; 10.

DR PROSITE; PS00230; MAP1B NEURAXIN; 6.

KW Microtubules; Repeat; Phosphorylation.

FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.

FT REPEAT 1878 1894 MAP1B 1.

FT REPEAT 1895 1911 MAP1B 2.

FT REPEAT 1912 1928 MAP1B 3.

FT REPEAT 1929 1945 MAP1B 4.

FT REPEAT 1946 1962 MAP1B 5.

FT REPEAT 1963 1979 MAP1B 6.

FT REPEAT 1997 2013 MAP1B 7.

FT REPEAT 2014 2030 MAP1B 8.

FT REPEAT 2031 2047 MAP1B 9.

FT REPEAT 2048 2064 MAP1B 10.

FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY KKEE AND KKEI/V REPEATS).

SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;

Query Match 11.6%; Score 92.5; DB 1; Length 2468;

Best Local Similarity 31.0%; Pred. No. 35;

Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

Qy 68 LSKDTFIKPVFKKIEKKKEENKPTFDVSKKONPQVNHSLQSHRKLQRE-----E 122

Db 638 VKKETVKP-----EDKKEKEKPKKEVAKEDKTPi---KKEKPKKEVKKEVKEIK 689

Qy 123 HSKQSDSTKDV 133

Db 690 KEEKKEPKKEV 700

RESULT 5

SPT7 YEAST

ID SPT7 YEAST STANDARD; PRT; 1332 AA.

AC P35177;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Transcriptional activator SPT7.

GN SPT7 OR YBR081C OR YBR0739.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=95229044; PubMed=7713415;

RA Gansheroff L.J., Dollard C., Tan P., Winston F.;

RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein important for transcription in vivo.";

RL Genetics 139:523-536(1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C;

RX MEDLINE=95076715; PubMed=7985423;

RA van der Aart Q.J.M., Barthe C., Doignon F., Aigle M., Crouzet M., Steensma H.V.;

RT "Sequence analysis of a 31 kb DNA fragment from the right arm of Saccharomyces cerevisiae chromosome II.";

RL Yeast 10:959-964(1994).

RN [3]

RP SEQUENCE OF 1-835 FROM N.A.

RC STRAIN=S288C;

RA Andre B., Criepluch C., Hein C., Jauniaux J.C., Urrestazu A., Vissers S.;

RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE OF 463-523 FROM N.A.

RX MEDLINE=92285152; PubMed=1350857;

RA Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J., David I.B.;

RT "The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins.";

RL Nucleic Acids Res. 20:2603-2603(1992).

CC -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY OTHER GENES.

CC -!- SUBCELLULAR LOCATION: Nuclear.

CC -!- SIMILARITY: Contains 1 bromodomain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AP003363; BAB57837.1; -  
 DR EMBL; AP003134; BAB42766.1; -  
 DR EMBL; AP004827; BAB95484.1; -  
 DR PIR; A89951; A89951.  
 DR HAMAP; MF\_00303; -; 1.  
 DR InterPro; IPR001179; FKBP\_PPIASE.  
 DR InterPro; IPR005215; Trig\_fac.  
 DR Pfam; PF00254; FKBP; 1.  
 DR TIGRFAMs; TIGR00115; tlg; 1.  
 DR PROSITE; PS00453; FKBP\_PPIASE\_1; FALSE\_NEG.  
 DR PROSITE; PS00454; FKBP\_PPIASE\_2; FALSE\_NEG.  
 DR PROSITE; PS00509; FKBP\_PPIASE\_3; 1.  
 DR Cell division; Chapterone; Isomerase; Rotamase; Complete proteome.  
 FT DOMAIN 163 248 PPIASE, FKBP-TYPE.  
 SQ SEQUENCE 433 AA; 48609 MW; 8865D9AP6A18C1E7 CRC64;

Query Match 11.2%; Score 89.5; DB 1; Length 433;  
 Best Local Similarity 22.1%; Pred. No. 9.1;  
 Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;  
 Yy 1 DTGVESELKPHRVTVT-IQNGKE--MSTIVSEEDFILPVYKG-ELEK----- 44  
 Db 81 DETDIKPEVAQPEVSVTQIEKGDFEATVVEPEVKGIDYKLEIEKQETLSDDLEQ 140  
 Yy 45 -----GYPDQWEISG--PEGKDGAGVNLNKTFTKP 76  
 Db 141 AIDSLGHLAMVVKEDGVNGDTVNDIFSG-SVDGEFEGQAGYDLEIGSGSFI-P 198  
 Yy 77 VFK-----KLEEKKE-----BE---NRPTFDVS---KKDNPNVHNSQLNE- 111  
 Db 199 GFEEQLGEMKVDKEEKVVVTPPEPVEHAEELAGKATFKTKVNEIKFEVPELTDIEANEL 258  
 Yy 112 -----SHRKEDLOREHSQKSDSTKDTATVLDKNISKSTTN 150  
 Db 259 DAEANTVDYKELNKRKLAEQKATDAENV-----EKEEAITKATDN 299

RESULT 8  
 YD2 SCHPO STANDARD; PRT; 1888 AA.  
 AC O14207;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C6B12.02c in chromosome I.  
 GN SPAC6B12.02C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule K., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,

RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Goffeau F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,  
 RA Cerrutti L., Lowe I., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Sipakowski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; Z98531; CAB11064.1; -  
 DR PIR; T39009; T39009.  
 DR GeneDB Spombe; SPAC6B12.02c; -  
 KW Hypothetical protein; Transmembrane.  
 FT TRANSMEM 697 717 POTENTIAL.  
 FT TRANSMEM 755 775 POTENTIAL.  
 FT TRANSMEM 866 886 POTENTIAL.  
 FT TRANSMEM 915 935 POTENTIAL.  
 FT TRANSMEM 977 997 POTENTIAL.  
 FT TRANSMEM 1033 1053 POTENTIAL.  
 FT TRANSMEM 1336 1356 POTENTIAL.  
 FT TRANSMEM 1645 1665 POTENTIAL.  
 FT DOMAIN -1662 1665 POLY-LEU.  
 SQ SEQUENCE 1888 AA; 217432 MW; 8AD3BBC32397C29 CRC64;

Query Match 11.1%; Score 89; DB 1; Length 1888;  
 Best Local Similarity 23.9%; Pred. No. 48;  
 Matches 37; Conservative 29; Mismatches 59; Indels 30; Gaps 7;

Yy 24 SSTIVSEEDF-----ILPVY---KGELEKGYQFDGWEISGFE-----GKDKAG 63  
 Db 390 SSSLTSENPFQNLVNAANAVSTIPVYRTTKTKKKN-RFKYVEVEKLPDLILESCKKAPK 448  
 Yy 64 YVNLNKTFTIKPVFKKIEEKKEENKPTFDVSKKDNPNVHNSQLNESHKEDLOREH 123  
 Db 449 FLRVFARSSSHIP--KMIRKQWDSKKYFDFKESDRQVLDVLSDWYSGKHELVOQSH 506  
 Yy 124 SQKSDS-TKDVATVLDKN-----NISKSTTN 151  
 Db 507 SYKKPSDSKSVGGNIFSVNSKKGSHVNINAKTANN 541

RESULT 9  
 DACA BACSU STANDARD; PRT; 443 AA.  
 AC P08750;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE D-alanyl-D-alanine carboxypeptidase precursor (EC 3.4.16.4) (DD-  
 DE peptidase) (DD-carboxypeptidase) (CPase) (PBPs).  
 GN DACA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;

RX MEDLINE=96051385; PubMed=7584024;  
RA Ogasawara N., Nakai S., Yoshikawa H.;  
RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
RT subtilis chromosome containing the replication origin.";  
RL DNA Res. 1:1-14(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunat F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Bruschi C., Caldwel B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Goughly E.J., Grandi G.,  
RA Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kashara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinou S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Neone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Forwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche R., Rose M., Sadaie Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,  
RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler H., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis.";  
RL Nature 390:249-256(1997).  
RN [3]  
RP SEQUENCE OF 32-102.  
RX MEDLINE=80182289; PubMed=6768745;  
RA Waxman D.J., Strominger J.L.;  
RT "Sequence of active site peptides from the penicillin-sensitive D-  
RT alanine carboxypeptidase of Bacillus subtilis. Mechanism of  
RT penicillin action and sequence homology to beta-lactamases.";  
RL J. Biol. Chem. 255:3964-3976(1980).  
RN [4]  
RP SEQUENCE OF 103-443 FROM N.A.  
RX MEDLINE=86250602; PubMed=3087956;  
RA Todd J.A., Roberts A.N., Johnstone K., Piggot P.J., Winter G.,  
RA Ellar D.J.;  
RT "Reduced heat resistance of mutant spores after cloning and  
RT mutagenesis of the Bacillus subtilis gene encoding penicillin-binding  
RT protein 5.";  
RL J. Bacteriol. 167:257-264(1986).  
RN [5]  
RP SEQUENCE OF 414-443.  
RX MEDLINE=81117303; PubMed=6780559;  
RA Waxman D.J., Strominger J.L.;  
RT "Primary structure of the COOH-terminal membranous segment of a  
RT penicillin-sensitive enzyme purified from two Bacilli.";  
RL J. Biol. Chem. 256:2067-2077(1981).  
CC -!- FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE  
CC CELL WALL PRECURSORS.  
CC -!- CATALYTIC ACTIVITY: D-alanyl-D-alanine + H(2)O = 2 D-alanine.  
CC -!- PATHWAY: Peptidoglycan synthesis; final stages.  
CC -!- SUBCELLULAR LOCATION: Membrane-associated.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11; ALSO KNOWN AS THE  
CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 1 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D26185; BAA05246.1; -;  
DR EMBL; Z99104; CAB11786.1; -;  
DR EMBL; M13766; AAA22375.1; -;  
DR PIR; S66040; S66040.  
DR MEROPS; S11.001; -;  
DR SubtilList; BG10074; dacA.  
DR InterPro; IPR001967; Ala/AlaCtpase1.  
DR Pfam; PF00768; Peptidase\_S11; 1.  
DR PRINTS; PR00725; DADACBPTASE1.  
KW Hydrolase; Carboxypeptidase; Peptidoglycan synthesis; Cell wall;  
KW Membrane; Signal; Complete proteome.  
FT SIGNAL 1 31  
FT CHAIN 32 443 D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.  
FT ACT\_SITE 67 67 ACYLATED BY PENICILLIN.  
FT CONFLICT 100 100 E -> Q (IN REF. 3).  
FT CONFLICT 227 227 E -> Q (IN REF. 4).  
SQ SEQUENCE 443 AA; 48636 MW; DA6C5B0307D7C117 CRC64;  
  
Query Match 11.1%; Score 88.5; DB 1; Length 443;  
Best Local Similarity 24.8%; Pred. No. 11;  
Matches 31; Conservative 23; Mismatches 44; Indels 27; Gaps 4;  
  
QY 4 EVSELKPRVTVTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEIFSGEKDAG 63  
DB 318 EGDQVKGHK-TISVDKGEKEVGIVTKAFSLPVKNGE-EKNYKAK----- 361  
QY 64 YVINLSKDTFTKVPFKKIEEKEENKFTFDVSKKQPNVHNSHNRKEDLQREH 123  
DB 362 --VTLNKDLNTPAVKKGKTKVGK-----LTAETGTGDKYGLNSDLGVDLVTKEN 410  
QY 124 SQKSD 128  
DB 411 VEKAN 415  
  
RESULT 10  
IGA2\_HAEIN STANDARD; PRT; 1702 AA.  
AC P45384;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).  
GN IGA.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HK715 / Serotype B;  
RX MEDLINE=92234949; PubMed=1373717;  
RA Poulsen K., Reinholdt J., Kilian M.;  
RT "A comparative genetic study of serologically distinct Haemophilus  
RT influenzae type 1 immunoglobulin A1 proteases.";  
RL J. Bacteriol. 174:2913-2921(1992).  
CC -!- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A  
CC PRODUCING INTACT FC AND FAB FRAGMENTS.  
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
CC certain Pro-I-xaa bonds in the hinge region. No small molecule  
CC substrates are known.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY





CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.  
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS  
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
 CC STABILIZING MICROTUBULES.  
 CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
 CC WITH MAP1A AND MAP1B PROTEINS.  
 CC -!- DOMAIN: Has a highly basic region with many copies of the sequence  
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is  
 CC responsible for the binding of MAP1B to microtubules.  
 CC -!- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED  
 CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH  
 CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION  
 CC OF MAP1B.  
 CC -!- SIMILARITY: TO MAP1A.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X51396; CAA35761.1; --  
 CC PIR; S07549; QRMSP1.  
 CC MGD; MGI:1306778; Mtap1b.  
 CC GO; GO:0016358; P:dendrite morphogenesis; IMP.  
 CC GO; GO:0001578; P:microtubule bundling; IMP.  
 CC InterPro: IPR000102; MAP1B\_neuraxin.  
 CC Pfam; PF00414; MAP1B\_neuraxin; 10.  
 CC PROSITE; PS00230; MAP1B\_NEURAXIN; 7.  
 CC Microtubules; Repeat; Phosphorylation.  
 CC CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.  
 CC REPEAT 1874 1890 MAP1B 1.  
 CC REPEAT 1891 1907 MAP1B 2.  
 CC REPEAT 1908 1924 MAP1B 3.  
 CC REPEAT 1925 1941 MAP1B 4.  
 CC REPEAT 1942 1958 MAP1B 5.  
 CC REPEAT 1959 1975 MAP1B 6.  
 CC REPEAT 1993 2009 MAP1B 7.  
 CC REPEAT 2010 2026 MAP1B 8.  
 CC REPEAT 2027 2043 MAP1B 9.  
 CC REPEAT 2044 2060 MAP1B 10.  
 CC DOMAIN 589 787  
 CC  
 CC SEQUENCE 2464 AA; 270408 MW; PBD3DD99CFDDBA87 CRC64;  
 CC  
 CC Query Match 10.9%; Score 87; DB 1; Length 2464;  
 CC Best Local Similarity 28.4%; Pred. No. 89;  
 CC Matches 29; Conservative 16; Mismatches 37; Indels 20; Gaps 4;  
 CC  
 CC QY 68 LSKDTFKVFK-KIEEKEENKPTFDYKKKNDPQVNHSQLNESHKEDLORE----- 121  
 CC Db 632 VTNDKVVKKKEIKTKLEKEE---KPKKEVVKKEDKTPL---KXDEKPRKEEKKKEIKKEI 686  
 CC QY 122 -----EHSQKSDSTKDVATVLDDKNNSKSTNNPNK 154  
 CC Db 687 KKEERKELKKEVKKEKTPDKAKKEVKKEEKKKEEKKK 728  
 CC  
 CC ID RAT1 YEAST STANDARD; PRT; 1006 AA.  
 CC AC Q02782;  
 CC DT 01-OCT-1993 (Rel. 27, Created)  
 CC DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 CC DE Ribonucleic acid trafficking protein 1 (5'-3' exoribonuclease)  
 CC DE (EC 3.1.11.-) (P116).  
 CC GN RAT1 OR HKEI OR TAP1 OR YOR048C.  
 CC OS Saccharomyces cerevisiae (Baker's yeast).

CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CC NCBI\_TaxID=4932;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE=92331925; PubMed=1628825;  
 CC RA Amberg D.C., Goldstein A.L., Cole C.N.;  
 CC "Isolation and characterization of RAT1: an essential gene of  
 CC Saccharomyces cerevisiae required for the efficient nucleocytoplasmic  
 CC trafficking of mRNA.";  
 CC Genes Dev. 6:1173-1189 (1992).  
 CC [2]  
 CC SEQUENCE FROM N.A., AND EXORIBONUCLEASE ACTIVITY.  
 CC MEDLINE=93109318; PubMed=8417335;  
 CC RA Kenna M., Stevens A., McCammon M., Douglas M.G.;  
 CC "An essential yeast gene with homology to the exonuclease-encoding  
 CC XRN1/KEM1 gene also encodes a protein with exoribonuclease  
 CC activity.";  
 CC Mol. Cell. Biol. 13:341-350 (1993).  
 CC [3]  
 CC SEQUENCE FROM N.A.  
 CC MEDLINE=93268292; PubMed=8497260;  
 CC RA Aldrich T.L., di Segni G., McConaughy B.L., Keen N.J., Whelen S.,  
 CC Hall B.D.;  
 CC "Structure of the yeast TAP1 protein: dependence of transcription  
 CC activation on the DNA context of the target gene.";  
 CC Mol. Cell. Biol. 13:3434-3444 (1993).  
 CC [4]  
 CC SEQUENCE FROM N.A.  
 CC Landt O., Hiesel R., Unsel M.;  
 CC Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 CC [5]  
 CC SEQUENCE FROM N.A.  
 CC Bohn C., Bolotin-Fukuhara M., Daigman-Fornier B., Dang D.V.,  
 CC Valens M.;  
 CC Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: May function in the processing and/or trafficking of  
 CC nuclear mRNA. May be involved in general transcription as well.  
 CC Possesses 5'->3' exoribonuclease activity degrading poly(A) to  
 CC mainly 5'-AMP.  
 CC -!- COFACTOR: Requires magnesium.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: Belongs to the 5'-3' exonuclease family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; S61567; AAB26818.1; --  
 CC DB EMBL; M95626; AAA34960.1; --  
 CC DB EMBL; L06011; AAA16950.1; --  
 CC DB EMBL; Z11746; --; NOT ANNOTATED\_CDS.  
 CC DB EMBL; Z74956; CAA99240.1; --  
 CC PIR; S20126; S20126.  
 CC SGD; S0005574; RAT1.  
 CC GO; GO:0005634; C:nucleus; IDA.  
 CC GO; GO:0004534; P:5'-3' exoribonuclease activity; IDA.  
 CC GO; GO:0006365; P:35S primary transcript processing; IMP.  
 CC GO; GO:0006396; P:RNA processing; IMP.  
 CC InterPro; IPR004859; Put 53exo.  
 CC Pfam; PF03159; XRN N; 1.  
 CC Nuclear protein; Hydrolase; Nuclease; Exonuclease; Repeat.  
 CC DOMAIN 525 528 POLY-GLU.  
 CC FT DOMAIN 793 797 POLY-ASN.  
 CC FT DOMAIN 955 999 CONTAINS 2 X SYD, 2 X NNY, AND 2 X  
 CC Y->H; IN ALLELE TAP1-1; ACTIVATES  
 CC TRANSCRIPTION OF THE PROMOTER-DEFECTIVE  
 CC YEAST SUP4 TRNA (TYR) ALLELE SUP4A53T61.  
 CC FT MUTAGEN 683 683  
 CC FT

```
SQ SEQUENCE 1006 AA; 115933 MW; SDD5B0245F3E12A CRC64;
Query Match
Best Local Similarity 10.8%; Score 86.5; DB 1; Length 1006;
Matches 32; Conservative 40; Mismatches 47; Indels 49; Gaps 8;

Qy 17 IONKEMSTIVBEDFILPVY--KGELEKGYQDFGWEISGFEKGKDAGYVINLSKDTFI 74
Db 415 MSKGDRHPTVATEQ---LQWYDTQGNLAKG---SWNLTTSD-----MVRLKKEML 460
Qy 75 -----KPVFKIEEKEENKPTEDVSKKDNPNQVHSHQ-----LNESHK 115
Db 461 ANEGNEEAIAKQOQSDKNMIDKSEIDDAVSKANKTNFNLAENVMKQKIINKHRL 520
Qy 116 EDLQREEH-----SQKSDSTKDVTTATVLDK-----NNISSKSTT 149
Db 521 EKDNEEEIAKSKVKTEKAESCDLDAEIKDEIVADVNDRENSETT 568

RESULT 14
MLP1 YEAST
ID MLP1 YEAST STANDARD; PRT; 1875 AA.
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MPL1".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L01992; AAA34783.1; -
CC EMBL; X73541; CA851948.1; -
CC EMBL; Z28320; CAA82174.1; -
CC PIR; S38173; S38173.
CC SGD; S0001803; MLP1.
CC GO; GO:0005635; C:nuclear membrane; IDA.
CC GO; GO:0005654; C:nucleoplasm; IDA.
CC GO; GO:0006606; P:protein-nucleus import; IDA.
CC Coiled coil; DNA repair.
KW DOMAIN 69 487 COILED COIL (POTENTIAL).
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).

SQ SEQUENCE 1006 AA; 115933 MW; SDD5B0245F3E12A CRC64;
Query Match
Best Local Similarity 10.8%; Score 86.5; DB 1; Length 1006;
Matches 32; Conservative 40; Mismatches 47; Indels 49; Gaps 8;

Qy 17 IONKEMSTIVBEDFILPVY--KGELEKGYQDFGWEISGFEKGKDAGYVINLSKDTFI 74
Db 415 MSKGDRHPTVATEQ---LQWYDTQGNLAKG---SWNLTTSD-----MVRLKKEML 460
Qy 75 -----KPVFKIEEKEENKPTEDVSKKDNPNQVHSHQ-----LNESHK 115
Db 461 ANEGNEEAIAKQOQSDKNMIDKSEIDDAVSKANKTNFNLAENVMKQKIINKHRL 520
Qy 116 EDLQREEH-----SQKSDSTKDVTTATVLDK-----NNISSKSTT 149
Db 521 EKDNEEEIAKSKVKTEKAESCDLDAEIKDEIVADVNDRENSETT 568

RESULT 15
GLNA PYRKO
ID GLNA PYRKO STANDARD; PRT; 443 AA.
AC O08457;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
GN GLNA.
OS Pyrococcus kodakaraensis.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Thermococcus.
OX NCBI_TaxID=69014;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOD1;
RX MEDLINE=97316461; PubMed=9172372;
RA Adul Rahman R.N.Z., Jongsareejit B., Fujiwara S., Imanaka T.;
RT "Characterization of recombinant glutamine synthetase from the
RT hyperthermophilic archaeon Pyrococcus sp. strain KOD1.";
RL Appl. Environ. Microbiol. 63:2472-2476(1997).
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D86222; BAA20530.1; -
CC HSSP; P06201; ILGR.
CC InterPro; IPR001691; GLN_synth.
CC InterPro; IPR004809; GLNA.
CC InterPro; IPR001637; GLNA_adenlytn.
CC Pfam; PF00120; gln-synt; 1.
CC Pfam; PF03951; gln-synt; 1.
CC ProDom; PD001057; Gln_synt_C; 1.
CC TIGRFAMs; TIGR00653; GlNA; 1.
CC PROSITE; PS00180; GLNA.1; 1.
CC PROSITE; PS00181; GLNA_ATP; 1.
CC Ligase.
KW BINDING. 362 362 AMP (UNDER CONDITIONS OF ABUNDANT
FT BINDING. 443 AA; 50259 MW; 9426DCCFEEF18168 CRC64;
SQ SEQUENCE 443 AA; 50259 MW; 9426DCCFEEF18168 CRC64;
```

Query Match 10.8%; Score 86; DB 1; Length 443;  
Best Local Similarity 17.8%; Pred. No. 17;  
Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;  
QY 38 YKLEKGYQFDGWEISGEGKADAGYVNLKDTFI-----74  
DB 42 YEAEVDGVSFGSSIPGEGIEDSLIFKADPSTVAETPWEGIGRVYIYKGBEPYQA 101  
QY 75 --KPVFKKIEEKEEN-----KPTFDVSKKKD-----100  
DB 102 DPGILKRVLEKLEKGLAKHIGPEPEFYFKNGTWELHPDSSGYPDLVGLDKAREIR 161  
QY 101 -----NPNVNSHLSHRKEDLQREHSQKSD---STKDVATATVLD 139  
DB 162 REIALYMPYIGLKPEVLHHEVGAKAHEIDFRYDEALRTADNIVSFKHVVKAAVE 215

## RESULT 16

MDN1\_HUMAN STANDARD; PRT; 5596 AA.  
ID Q9NU22; O15019;  
AC 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Midasin (MIDAS-containing protein).  
GN MDN1 OR KIAA0301.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RX PubMed=12102729;  
RA Garbarino J.E., Gibbons I.R.;  
RT "Expression and genomic analysis of midasin, a novel and highly  
conserved AAA protein distantly related to dynein.";  
RL BMC Genomics 3:18-18(2002).  
RN [2]  
RP SEQUENCE OF 1255-2356 AND 3550-5596 FROM N.A.  
RC TISSUE=Testis;  
RA Tracey A.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 3550-5596 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97349984; PubMed=9205841;  
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. VII.  
The complete sequences of 100 new cDNA clones from brain which can  
code for large proteins in vitro.";  
RL DNA Res. 4:141-150(1997).  
CC -!- FUNCTION: May function as a nuclear chaperone and be involved in  
nucleus.  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- SIMILARITY: Contains 1 WWFA domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF503925; AAM7722.1; -  
CC EMBL; AL096678; CAB86660.1; -  
CC EMBL; AL096678; CAB86661.1; -  
CC EMBL; AB002299; BAA20761.1; -  
CC Genew; HGNC:18302; MDN1.

DR GO; GO:0005634; C:nucleus; NAS  
DR GO; GO:0016887; F:ATPase activity; NAS.  
DR GO; GO:0003754; F:chaperone complex activity; NAS.  
DR GO; GO:0006461; F:protein complex assembly; NAS.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR002035; WWFA\_A.  
DR Pfam; PF00004; AAA; 1.  
DR SMART; SM00382; AAA; 7.  
DR PROSITE; PS00327; WWFA; 1.  
DR PROSITE; PS02334; WWFA; 1.  
KW Chaperone; ATP-binding; Repeat; Nuclear protein.  
FT NP\_BIND 329 336 ATP (POTENTIAL).  
FT NP\_BIND 677 684 ATP (POTENTIAL).  
FT NP\_BIND 1084 1091 ATP (POTENTIAL).  
FT NP\_BIND 1390 1397 ATP (POTENTIAL).  
FT NP\_BIND 1753 1760 ATP (POTENTIAL).  
FT NP\_BIND 2066 2073 ATP (POTENTIAL).  
FT DOMAIN 3566 3573 POLY-GLU.  
FT DOMAIN 4784 4791 POLY-GLU.  
FT DOMAIN 5008 5013 POLY-GLU.  
FT DOMAIN 5182 5187 POLY-GLU.  
FT DOMAIN 5384 5583 WWFA.  
FT CONFLICT 2287 2312 RLFLSMDPVHGDISRMRNRLGIYI -> S (IN REF.  
SQ SEQUENCE 5596 AA; 632802 MW; 586C626161F96D4 CRC64;

Query Match 10.8%; Score 86; DB 1; Length 5596;  
Best Local Similarity 23.4%; Pred. No. 2.5e+02;  
Matches 30; Conservative 27; Mismatches 63; Indels 8; Gaps 2;  
QY 8 LKPHRVTVTIQNGKMSSTIVSEEDFILPVVYKGELEK-----YQFDGWEISGEGKDD 61  
DB 4629 LATHRSTAKLSVLAQVETELAQKGFCLPKFEMDSAGEGATEFHDYEGGIGEGEKMD 4688  
QY 62 AGYVNLKSDTFIKPVFKKIEEKEENKPTFDVSKKDNPNVNSHLSHRKEDLORE 121  
DB 4689 VS--DQIGNEQVEDTFQKGQEKEDPDSKDKINGEDNAIEMSEDFQKWHGDELEGE 4745  
QY 122 EHSQKSDS 129  
DB 4747 EDDEKSDS 4754

## RESULT 17

BPBA\_BACSU STANDARD; PRT; 914 AA.  
ID BPBA\_BACSU  
AC P39793;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Penicillin-binding protein 1A/1B (PBP1) [Includes: Penicillin-  
insensitive transglycosylase (EC 2.4.2.-) (Peptidoglycan TGase);  
Penicillin-sensitive transpeptidase (EC 3.4.-.-) (DD-transpeptidase)].  
GN PONA.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 499-515.  
RC STRAIN=168;  
RX MEDLINE=95113769; PubMed=7814321;  
RA Popham D.L., Setlow P.;  
RT "Cloning, nucleotide sequence, and mutagenesis of the Bacillus  
subtilis ponA operon, which codes for penicillin-binding protein  
(PBP) 1 and a PBP-related factor.";  
RL J. Bacteriol. 177:326-335(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168 / Marburg;  
RX MEDLINE=96349105; PubMed=8760912;  
RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,  
RA Serron P.;  
RT "Sequence analysis of the Bacillus subtilis chromosome region between



OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
RC STRAIN=S288c; PubMed=8657111;  
RX MEDLINE=96239492; PubMed=8657111;  
RA Chen G.-C., Zheng L., Chan C.S.M.;  
RT "The LIM domain-containing Dbp1 GTPase-activating protein is required  
RT for normal cellular morphogenesis in *Saccharomyces cerevisiae*.";  
RL Mol. Cell. Biol. 16:1376-1390(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c / FY1679;  
RX MEDLINE=97060020; PubMed=8904341;  
RA Wiemann S., Rechmann S., Benes V., Voss H., Schwager C., Vlcek C.,  
RA Stegemann J., Zimmermann J., Erfle H., Paces V., Ansoerge W.;  
RT "Sequencing and analysis of 51 kb on the right arm of Chromosome XV  
RT from *Saccharomyces cerevisiae* reveals 30 open reading frames.";  
RL Yeast 12:281-288(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97344368; PubMed=9200815;  
RA Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,  
RA Schwager C., Paces V., Sander C., Ansoerge W.;  
RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV.";  
RL Yeast 13:655-672(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c;  
RX MEDLINE=96101594; PubMed=7498791;  
RA Stevenson B.J., Ferguson B., de Virgilio C., Bi E., Pringle J.R.,  
RA Ammerer G., Sprague G.F. Jr.;  
RT "Mutation of RGA1, which encodes a putative GTPase-activating protein  
RT for the polarity-establishment protein Cdc42p, activates the  
RT pheromone-response pathway in the yeast *Saccharomyces cerevisiae*.";  
RL Genes Dev. 9:2949-2963(1995).  
RN [5]  
RP SEQUENCE OF 570-639 FROM N.A.  
RC STRAIN=SNY243;  
RX MEDLINE=93087574; PubMed=1454852;  
RA Famer S.W., Elledge S.J., Davis R.W.;  
RT "Dominant genetics using a yeast genomic library under the control of  
RT a strong inducible promoter.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:11589-11593(1992).  
CC -1- FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR CDC42 AND/OR RHO1.  
CC NEGATIVE REGULATOR OF THE PHEROMONE-RESPONSE PATHWAY THROUGH THE  
CC STE20 PROTEIN KINASE. ACTS AT A STEP BETWEEN THE G-PROTEIN AND THE  
CC MAP KINASE MODULE. DOMINANT SUPPRESSOR OF BUD EMERGENCE DEFECT  
CC CAUSED BY DELETION OF IPL2/BEM2. INVOLVED IN THE CONTROL OF  
CC POLARIZED CELL GROWTH AND PROPER BUD SITE SELECTION.  
CC -1- SIMILARITY: Contains 2 LIM zinc-binding domains.  
CC -1- SIMILARITY: Contains 1 Rho-GAP domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U07421; AAA16875.1; -  
CC EMBL; X90518; CAA62108.1; -  
CC EMBL; X94335; CAA64046.1; -  
CC EMBL; Z75035; CAA9326.1; -  
CC EMBL; X90950; CAA62445.1; -  
CC EMBL; L02617; AAA35153.1; -  
CC PIR; S48535; S48535.  
CC SGD; S0005653; RGA1.  
CC GO; GO:0005100; F:Rho GTPase activator activity; IPI.  
CC GO; GO:0004871; F:signal transducer activity; IPI.  
CC GO; GO:0007015; P:actin filament organization; IMP.

DR GO; GO:0007118; P:apical bud growth; IPI.  
DR GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. . .; IPI.  
DR GO; GO:0007125; P:invasive growth; IPI.  
DR GO; GO:0007119; P:isotropic bud growth; IPI.  
DR GO; GO:0007124; P:pseudohyphal growth; IPI.  
DR GO; GO:0000750; P:signal transduction during conjugation with. . .; IGI.  
DR GO; GO:0007264; P:small GTPase mediated signal transduction; IPI.  
DR InterPro; IPR001781; LIM.  
DR InterPro; IPR00198; RhoGAP.  
DR Pfam; PF00412; LIM; 1.  
DR Pfam; PF00620; RhoGAP; 1.  
DR ProDom; PD000094; LIM; 2.  
DR SMART; SM00132; LIM; 2.  
DR SMART; SM00324; RhoGAP; 1.  
DR PROSITE; PS00478; LIM DOMAIN\_1; 1.  
DR PROSITE; PS50023; LIM DOMAIN\_2; 2.  
DR PROSITE; PS50238; RHO GAP; 1.  
KW GTPase activation; Repeat; LIM domain; Metal-binding; Zinc;  
KW Pheromone response.  
FT DOMAIN 13 66 LIM 1.  
FT DOMAIN 70 122 LIM 2.  
FT DOMAIN 791 1006 RHO-GAP.  
FT VARIANT 866 866 V -> A.  
FT VARIANT 898 898 K -> R.  
FT VARIANT 926 926 S -> G.  
FT MUTAGEN 40 40 C->S: BIPOLAR BUDDING.  
FT MUTAGEN 37 37 C->S: BIPOLAR BUDDING.  
FT MUTAGEN 98 98 C->S: BIPOLAR BUDDING.  
FT MUTAGEN 101 101 C->S: BIPOLAR BUDDING.  
FT CONFLICT 457 457 D -> E (IN REF. 4).  
FT CONFLICT 507 507 T -> P (IN REF. 4).  
SQ SEQUENCE 1007 AA; 112831 MW; C805411B57553791 CRC64;  
Query Match 10.7%; Score 85.5; DB 1; Length 1007;  
Best Local Similarity 21.0%; Pred. No. 44;  
Matches 34; Conservative 34; Mismatches 69; Indels 25; Gaps 3;  
QY 1 DTGEVSELKPHRVTVTQNGKEMSTI-----VSEDFILPVYGELEKGYQDFGWEIS 54  
DB 349 NTGEISQMDPSLSRKVLNNIVEETNALQRPVVVVKEDRSVPDLAGVQOQAEKYSYNN 408  
QY 55 GPEGKKDAGYVNLKSDTFIKPFVKKIEEKEENKPTFDVSKKDNPOVNHSQLNESH 114  
DB 409 SGKGRKISRSRRSRKDLMI-----NLKSRATKQDSNVKLSPAKVTSSR 454  
QY 115 KEDLQREHSQKSDSTKDVATVLD-----KNNISKSTTNN 151  
DB 455 SQDLARDNDSDHGLDTPNSNSTSLDLVNNQKSLNKRFTDN 496  
RESULT 19  
IF2 STAEP  
ID IF2 STAEP STANDARD; PRT; 720 AA.  
AC Q8CST4.  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Translation initiation factor IF-2.  
GN INFB OR SE0945.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
RA Chen Z., Wen Y.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBSJ databases.  
CC -1- FUNCTION: One of the essential components for the initiation of  
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous  
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.  
CC Also involved in the hydrolysis of GTP during the formation of the  
CC 70S ribosomal complex (By similarity).



```

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE016747; AA004542.1; -
DR HAMAP; MF_00100; -; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU D2; 1.
DR Pfam; PF04760; IF2_N_2; 1.
DR ProDom; PD186100; IF2; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS01176; IF2; 1.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 225 373 G-DOMAIN.
FT NP_BIND 231 238 GTP (BY SIMILARITY).
FT NP_BIND 277 281 GTP (BY SIMILARITY).
FT NP_BIND 331 334 GTP (BY SIMILARITY).
SQ SEQUENCE 720 AA; 79343 MW; 07FB5A6A59CF970C CRC64;
Query Match 10.6%; Score 85; DB 1; Length 720;
Best Local Similarity 28.0%; Pred. NO. 34;
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;
Oy 81 IEEKEEENKPTEDVSKKD---NPQVNSQLNESHRKEDLQREHSQKSDSTK---D 132
Db 38 LEEEOIKALDKKFKASQASQDKNTQNTQNNHOKSNKQNSDKKQSKNKPTKKEQN 97
Oy 133 VTATVLDKNNISSTNNPNK 154
Db 98 NKGQOQNNKNTKNQKNK 119
RESULT 20
ID UN89 CAEEL STANDARD; PRT; 6632 AA.
AC 001761; Q17362;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OX Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]_TaxID=6239;
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT assembly, encodes a giant modular protein composed of Ig and signal
RT transduction domains".
RL J. Cell Biol. 132:835-848 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Le T.T., Wilson R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofilament

```

```

CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U33058; AAB00542.1; -
DR EMBL; AF003131; AAB54132.2; -
DR PDB; 1FHO; 20-DEC-00.
DR WormPep; C09D1.1; CE30426.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001849; PH.
DR InterPro; IPR007850; RCSD.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; ig; 47.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF05177; RCSD; 5.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IG2; 23.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50010; DH 2; 1.
DR PROSITE; PS50835; IG LIKE; 49.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW 3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.

```



```
FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBRONECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3085 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT CONFLICT 2137 2137 A -> P (IN REF. 1).
FT CONFLICT 2245 2247 AKA -> PKP (IN REF. 1).
FT CONFLICT 2258 2258 A -> P (IN REF. 1).
FT CONFLICT 2284 2284 E -> G (IN REF. 1).
FT CONFLICT 2297 2297 M -> I (IN REF. 1).
FT CONFLICT 3531 3531 A -> G (IN REF. 1).
FT CONFLICT 3884 3888 DAGGY -> RRRRI (IN REF. 1).
FT CONFLICT 3929 3929 A -> V (IN REF. 1).
FT CONFLICT 5134 5134 A -> P (IN REF. 1).
FT CONFLICT 5145 5145 T -> S (IN REF. 1).
FT CONFLICT 5185 5185 G -> A (IN REF. 1).
FT CONFLICT 5199 5199 K -> N (IN REF. 1).
FT CONFLICT 5202 5202 L -> F (IN REF. 1).
FT CONFLICT 5213 5213 F -> L (IN REF. 1).
FT CONFLICT 6178 6178 A -> G (IN REF. 1).
FT CONFLICT 6268 6268 K -> E (IN REF. 1).
SQ SEQUENCE 6632 AA; 731665 MW; 262D3EDD62960E89 CRC64;

Query Match 10.6%; Score 85; DB 1; Length 6632;
Best Local Similarity 22.1%; Pred. No. 3.6e+02;
Matches 34; Conservative 32; Mismatches 68; Indels 20; Gaps 7;

QY 7 ELKPHRVTVTQNGKMSSTIVSEEDF--ILPVYKGELEKGYQFDGWE--ISGFEGKGD 61
DB 1143 DLRP--VSLTVKSGE---AVFSAHAFGLPLTYEWSVNGRKVRDQCGEARVTRDESTVD 1197
```

```
DR Pfam: PF03797; Autotransporter; 1.
DR Pfam: PF0395; IGA1; 1.
DR Pfam: PF03212; Pectactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR TIGFAMs; TIGF01414; autotrans_bar1; 1.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal;
Complete proteome.
KM SIGNAL
FT CHAIN 1 25
FT PROPEP 1015 1694 IMMUNOGLOBULIN A1 PROTEASE.
FT ACT SITE 288 288 HELPER PEPTIDE (POTENTIAL).
FT CONFLICT 253 254 PROBABLE.
FT CONFLICT 272 272 EN -> GV (IN REF. 1).
FT CONFLICT 444 444 G -> A (IN REF. 1).
FT CONFLICT 464 464 S -> E (IN REF. 1).
FT CONFLICT 866 866 S -> T (IN REF. 1).
FT CONFLICT 1036 1036 A -> D (IN REF. 1).
FT CONFLICT 1074 1074 A -> G (IN REF. 1).
FT CONFLICT 1421 1421 A -> G (IN REF. 1).
FT CONFLICT 1545 1545 H -> T (IN REF. 1).
SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;

Query Match 10.6%; Score 84.5; DB 1; Length 1694;
Best Local Similarity 26.1%; Pred. No. 91;
Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;

Qy 66 INLSKDTPIKPVFKKIEEKKENKPTFVSKKKNPQVNHQSLNESHKEDLQREHS- 124
Db 1288 INTGSAITITERAEKSDKQPTASTEDASQHKNTADVANSSESSDKSRRRRSI 1347
Qy 125 --QKSDSTKDVATVLDKNNISSKSTNNPK 154
Db 1348 SOPQTSAEETTAASTDETTITADNSKSKSPNR 1379

RESULT 22
YF8 YEAST STANDARD; PRT; 778 AA.
ID YF8 YEAST
AC P43610;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 88.7 kDa helicase in CDC26-SAP155 intergenic region.
GN YFR038W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268 (1995).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=96287654; PubMed=8666381;
RA Eki T., Naitou M., Hagiwara H., Abe M., Ozawa M., Sasanuma S.-I.,
RA Sasanuma M., Tsuchiya Y., Shibata T., Watanabe K., Ono A.,
RA Yamazaki M.-A., Tashiro H., Hanaoka F., Murakami Y.;
RT "Fifteen open reading frames in a 30.8 kb region of the right arm of
RT chromosome VI from Saccharomyces cerevisiae.";
RL Yeast 12:177-190 (1996).
RL [3]
RP SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC This SNF2-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D50617; BAA09277.1; -
DR PIR; S56293; S56293.
DR SCD; S0001934; YFR038W.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR00330; SNF2_N.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDc; 1.
DR SMART; SM00490; HELICc; 1.
KW Hypothetical protein; Nuclear protein; DNA-binding; Helicase;
KW ATP-binding.
FT NP BIND 247 254 ATP (POTENTIAL).
FT SITE 352 355 DEGH BOX.
SQ SEQUENCE 778 AA; 88730 MW; 3B6C08575EABD84 CRC64;

Query Match 10.5%; Score 84; DB 1; Length 778;
Best Local Similarity 25.5%; Pred. No. 43;
Matches 37; Conservative 18; Mismatches 56; Indels 34; Gaps 5;

Qy 23 MSSTVSEEDPILPVYKGELEKGYQFDGMEISGFEGKDGAGVYINLSKDTPIKPVFKKIE 82
Db 72 LDDVHSDEDI-----QLDSESDSTEAQV-AQVVDLADDT--KSKQSLD 115
Qy 83 EKKKEENKPTFVSKKKNPQVNHQSLNESHKEDLQREHS- 124
Db 116 DELSMPTKTVSLKIKLNEFVRSQVSSIIADTLHRSNEVANANTKONSNDDEHS 175
Qy 125 QKSDSTKDVATVLDKNNISSKSTNNPK 149
Db 176 SKKRRTKKKSIYDFPKKKKNNEDTT 200

RESULT 23
U2R1 HUMAN STANDARD; PRT; 479 AA.
ID U2R1 HUMAN
AC Q15655; Q13570;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit
DE related-protein 1.
GN U2AF1-RS1 OR U2AF1RS1 OR U2AFBPL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96163878; PubMed=8586425;
RA Kitegawa K., Wang X., Hataida I., Yamaoka T., Nojima H.,
RA Inazawa J., Abe T., Mitsuura K., Oshimura M., Murata A., Monden M.,
RA Mukai T.;
RT "Isolation and mapping of human homologues of an imprinted mouse gene
RT U2af1-rbl.";
RL Genomics 30:257-263 (1995).
RL [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96212931; PubMed=8630064;
RA Pearshall R.S., Shibata H., Brozowska A., Yoshino K., Okuda K.,
RA DeJong P.J., Plas C., Chapman V.M., Hayashizaki Y., Held W.A.;
RT "Absence of imprinting in U2AFBPL, a human homologue of the imprinted
RT mouse gene U2afp-rb.";
RL Biochem. Biophys. Res. Commun. 222:171-177 (1996).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -1- SIMILARITY: Contains 2 CCH1-type zinc fingers.
```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D49676; BAA08532.1; -
CC EMBL; U51224; AAA98669.1; -
CC MIM; 601075; -
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0003723; F:RNA binding activity; NAS.
CC InterPro; IPR000504; RNA_rec_mot.
CC InterPro; IPR000571; Znf_CCH.
CC Pfam; PF00076; rrm; 1.
CC Pfam; PF00642; zf-CCH; 2.
CC SMART; SM00356; ZNF_CCH1; 2.
CC SMART; PS50102; RRM; 1.
CC PROSITE; PS00030; RRM_RNP_1; FALSE NEG.
CC NUCLEAR protein; RNA-binding; Ribonucleoprotein; Zinc-finger; Repeat.
CC DOMAIN 203 309
CC SEQUENCE 479 AA; 57643 MW; 96F326694BD4E7C0 CRC64;

SQ
Query Match 10.5%; Score 83.5; DB 1; Length 479;
Best Local Similarity 21.7%; Pred. No. 28;
Matches 26; Conservative 25; Mismatches 40; Indels 29; Gaps 4;

QY 60 KDAGYVINKQFIKRVFKKIEKK-----BEKKPFDVSKKKDNQ 103
DB 44 RDSGLSGOEEDFTFIE--QGLEEKLERERERLHEWLLRQKQOEERIKKEBBA 101
QY 104 VN-----HSQLNESHKEDLQREHSHQSKSDTQVATVLD--KNNISKSTNNP 152
DB 102 KKLLEGEQRLTKQWQEQKQKEKEBEEBQKQEKKEKEENQKMLDQAEMLDSTNNQ 161

RESULT 24
AR52 DROME STANDARD; PRT; 943 AA.
ID AR52 DROME STANDARD; PRT; 943 AA.
AC 09V9K7;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arsenite-resistance protein 2 homolog.
GN CG7843.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10721132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle J.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer J.R., Champe W., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abirl J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Beeson K.Y., Basu P.V., Baxendale J., Bayraktoglu L., Beasley E.M.,
RA Beckson K.Y., Benos P.V., Berman J.P., Bhandari D., Bolshakov S.,
RA Borokov D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu L.B., Davies P.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Dou P.L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

```

```

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.H., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft K., Kravitz S., Kulp D., Lai Z.,
RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mikhlov G., Misha N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RC STRAIN=Berkeley;
RX MEDLINE=22426069; PubMed=12537572;
RA Miste S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktoglu L., Berman B.P.,
RA Beltencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
CC -1- ALTERNATIVE PRODUCTS.
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q9V9K7-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q9V9K7-2; Sequence=VSP 000327;
CC -1- SIMILARITY: BELONGS TO THE AR52 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE003784; AAM68343.1; -
CC EMBL; AE003784; AAM68345.1; -
CC FlyBase; FBgn0033062; CG7843.
CC Pfam; PF04959; AR52; 1.
CC Hypothetical protein; Alternative splicing.
CC VARSPLIC 47 50 Missing (in isoform short).
CC /FTId=VSP 000327.
CC SEQUENCE 943 AA; 107221 MW; 0CIAF05E02E8AB08 CRC64;

SQ
Query Match 10.5%; Score 83.5; DB 1; Length 943;
Best Local Similarity 25.0%; Pred. No. 58;
Matches 36; Conservative 23; Mismatches 68; Indels 17; Gaps 4;

QY 4 EYSELKP-----HRYTVITQNGKWSSTI-----VSEEDILPYKGELEKGYQFDGM 51
DB 272 KYLDEPKDPVYERAEQMSVKEKTTNSPKKESADYVSTQKRVYRPNVSGENW 331
QY 52 EISGEFGKDGAVINLSKQFIKRVFKKIE--KKEENKPTPDVSKKKDNPNQVNHQ 108
DB 332 DDDDAENSAKPKKELADSDSDSKPDQINKKTKKKRRNSDDSSSSSSSSSDEDEK 391

```

OY 109 LINESHREDLOREHSHOKSDSTKD 132  
DB 392 LKEKYDVEDGJRAE--QKTEAKD 413

RESULT 25

IF31\_HUMAN STANDARD; PRT; 258 AA.  
ID IF31\_HUMAN  
AC 075822; Q9BUD2; Q9H8Q2;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Eukaryotic translation initiation factor 3 subunit 1 (eIF-3 alpha)  
DE (eIF3 p35) (eIF3j).  
GN EIF3J.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=99041954; PubMed=9822659;  
RA Block K.L., Vornlocher H.-P., Hershey J.W.B.;  
RT "Characterization of cDNAs encoding the p44 and p35 subunits of human  
RL translation initiation factor eIF3";  
RL J. Biol. Chem. 273:31901-31908(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovarian carcinoma;  
RA Itooi T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,  
RA Wagaetsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo Y., Takiguchi S.,  
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,  
RA Niomiya K., Iwayanagi T.;  
RT "NEDO human cDNA sequencing project";  
RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=uterus;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buelow K.H., Scheffer C.F., Bat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueda T.B., Teshiyuki S., Carinci P., Prange C.,  
RA Rata S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Scherch A., Schin J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RL human and mouse cDNA sequences";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
CC -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF  
CC METHIONYL-TRNAI AND MRNA.  
CC -1- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 12 DIFFERENT SUBUNITS (By  
CC similarity).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@eb-sib.ch](mailto:license@eb-sib.ch).  
CC -----  
CC EMBL; U97670; AAC78729.1; -.  
CC DR EMBL; AK023388; BAB14555.1; -.  
CC DR EMBL; BC002719; AAH02719.1; -.  
CC DR Genem; HGNC:3270; EIF3J1.  
CC DR GK; 075822; -.  
CC DR MIM; 603910; -.  
CC DR GO; GO:0005852; C:eukaryotic translation initiation factor 3. . .; TMS.  
CC DR GO; GO:0003743; F:translation initiation factor activity; TMS.  
CC DR GO; GO:0006446; P:regulation of translational initiation; TMS.  
CC KM Initiation factor; Protein biosynthesis.  
CC FT DOMAIN 2 8 POLY-ALA.  
CC FT DOMAIN 29 32 POLY-GLY.  
CC FT DOMAIN 53 57 POLY-ASP.  
CC FT DOMAIN 218 224 POLY-LYS.  
CC FT CONFLICT 40 40 E -> G (IN REF. 1).  
CC FT CONFLICT 141 141 A -> T (IN REF. 2).  
CC SQ SEQUENCE 258 AA; 29062 MW; 8362423542445AA CRC64;  
OY Query Match 10.4%; Score 83; DB 1; Length 258;  
OY Best Local Similarity 25.4%; Freq. No. 16;  
OY Matches 30; Conservative 25; Mismatches 51; Indels 12; Gaps 4;  
DB 38 YKGELEKGYQPDGWEISGFEGKXDGY--VINLSKDTFIRPKKIEKEEENKPTPDV 95  
DB 39 WEGEDEEDVDKNDNDDDEKKEAEVPEVKISK--KIAEKIKERQKQKROEI 95  
OY 96 SKKDNPQ-----VNHSQINESHREDLOREHSHOKSDSTKYDTATV--LDKNITSK 146  
OY 96 KKRLEPEPEPVLPPEQLADKLRLKQLQESDLELAKETGVNNAVYGIDAMPSSR 153  
DB

RESULT 26

T0NB\_HAEIN STANDARD; PRT; 270 AA.  
ID T0NB\_HAEIN  
AC P42872;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE T0NB protein.  
GN T0NB OR H10251.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NTHI TN106;  
RX MEDLINE=94245357; PubMed=8188372;  
RA Jarosik G.P., Sanders J.D., Cope L.D., Muller-Eberhard U.,  
RA Hansen E.J.;  
RT "A functional t0nb gene is required for both utilization of heme and  
RL virulence expression by Haemophilus influenzae type b.";  
RL Infect. Immun. 62:2470-2477(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA Kennedy K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Wetman J.F., Phillips C.A., Spirigs T., Hedblom E., Cotton M.D.,  
RA Utecherback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geophagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RL Rd.";  
RL Science 269:496-512(1995).  
CC -1- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT  
CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO

[illegible]

```

RP SEQUENCE FROM N.A., AND SEQUENCE OF 303-315.
RC STAIN=9;
RX MEDLINE=93171870; PubMed=8436847;
RA Farrell A.W., Foster T.J., Holland K.T.;
RA "Molecular analysis and expression of the lipase of Staphylococcus
RT epidermidis";
RL J. Gen. Microbiol. 139:267-277 (1993).
RN [2]
RC SEQUENCE FROM N.A.
RC STAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Yu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
CC fatty acid anion.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS
CC NEGATIVELY REGULATED BY BACTERIOPHAGE LYSOGENIZATION (LIPASE
CC CONVERSION).
CC -1- SIMILARITY: BELONGS TO THE AHYDROLASE SUPERFAMILY. LIPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M95577; AAA19729.1; -.
CC EMBL; AE016744; AAC003878.1; -.
CC PIR; A47705; A47705.
CC InterPro; IPR005877; Gpos_YSTRK.
CC InterPro; IPR000734; Lipase_YSTRK.
CC InterPro; IPR000379; Ser_estr_site.
CC Pfam; PF04650; YSTRK_signal; 1.
CC TIGRFAMs; TIGR01168; YSTRK_signal; 1.
CC PROSITE; PS00120; LIPASE_SER; 1.
CC HydroLase; Lipid degradation; Zymogen; Signal; Complete proteome.
CC SIGNAL 1 35 REMOVED IN THE MATURE FORM.
CC PROPEP 36 302
CC CHAIN 303 688 LIPASE.
CC FT ACT_SITE 418 418 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 648 648 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT CONFLICT 96 96 W -> L (IN REF. 1).
CC FT CONFLICT 120 120 E -> G (IN REF. 1).
CC FT CONFLICT 120 120
CC SQ SEQUENCE 688 AA; 77343 MW; 6C95DB3A78AF86F6 CRC64;

Query Match 10.4%; Score 83; DB 1; Length 688;
Best Local Similarity 24.5%; Pred. No. 45;
Matches 35; Conservative 28; Mismatches 60; Indels 20; Gaps 6;

OY 16 TIQNGKEMSSITVEEEDFILPVYGELEKGYQFGWEISGFE-GKQDAGVYINLSKDTFI 74
DB 79 SITNESIHNHTPPKNEDMI-----QQQKDSQNKNKSESVVEQNKKEAFQNHSEF--- 129
OY 75 KPVFKIE-EKKEENKPTFVDSKKKNPOY-NHSQLNESHRREDIQREEHQSKSDSTK 131
DB 130 KPQGEQVELEKHAENNOTLHSKAAQSNEDVKTSPQLDNTAAKQEDSKENLTKQDTQS 189
OY 132 DVYATVLDKXNNISKSSTTNPNK 154
DB 190 SKTIDL-----RATRNOSK 205

RESULT 28
ID ADVSV_HUMAN STANDARD; PRT; 715 AA.
AC Q9Y6U3; O8MU97;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Adeverin (Scinderin).

```

GN SCIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OK NCBI\_TaxId=9606;  
 (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=22389257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.F., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Umed T.B., Ioshizuka S., Carminci P., Frange C.,  
 RA Bae S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Roach S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.:  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RL [2]  
 RP SEQUENCE OF 1-527 FROM N.A.  
 RL Kalicki J., Smith-Craig R.:  
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RN SEQUENCE OF 248-715 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Isozaki T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanabe T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arima M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.:  
 RT "NEDO human cDNA sequencing project.";  
 RT Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.  
 RL [1]  
 RL FUNCTION: Ca(2+)-dependent actin filament-severing protein that is  
 RL presumed to have a regulatory function in exocytosis by affecting  
 RL the organization of the microfilament network underneath the  
 RL plasma membrane. In vitro, also has barbed end capping and  
 RL nucleating activities in the presence of Ca(2+).  
 CC -1- SIMILARITY: BELONGS TO THE VILININ/GELSOLIN FAMILY.  
 CC -1- SIMILARITY: Contains 6 gelsolin-like repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: BC021090; AA021090.1; -;  
 DR EMBL: AC005281; AAD15423.1; -;  
 DR EMBL: AK027778; BAB55361.1; -;  
 DR HSSP: P02640; 2VIL.  
 DR InterPro: IPR001974; Gelsolin.  
 DR Pfam: PF00626; Gelsolin; 6.  
 DR PRINTS: PR00597; GELSOLIN.  
 DR SMART: SM00262; GEL, 6.  
 DR CyscoReceptor: Actin-binding; Repeat; Calcium; Capping protein.  
 KM DOMAIN 1 363  
 FT DOMAIN 364 715  
 FT REPEAT 27 76  
 FT REPEAT 148 188  
 FT REPEAT 265 307  
 FT GELSOLIN-LIKE 1.  
 FT GELSOLIN-LIKE 2.  
 FT GELSOLIN-LIKE 3.

FT REPEAT 398 451  
 FT REPEAT 523 564  
 FT REPEAT 626 668  
 FT SITE 112 119  
 FT SITE 138 146  
 FT SITE 146 146  
 FT CONFLICT 61 61  
 FT SEQUENCE 715 AA; 80508 MW; 45FB842C8CBFDB0 CRC64;  
 SQ  
 Query Match  
 Best Local Similarity 23.1%; Pred. No. 47;  
 Matches 34; Conservative 30; Mismatches 47; Indels 36; Gaps 5;  
 QY 12 RTVTIIOGKMSSTIVSEEDPILPVYKGELEKYGQFGWETSGPEGK-----KDAGV 65  
 DB 262 RTVVAENPNPSSMAWLESECHTLD--HGAQKIVWKGKDNPPQRAAKTAEEFLQ 319  
 QY 66 INLSKDFEIK-----PVFK-----IEEKKEENKPTDVSKKK 99  
 DB 320 MMYSKNTQIOVLPGEGETPIFKQFKDWRDKDQSGFGKVVYTEKVAQIKQIPFASKLH 379  
 QY 100 DNPQV--NHSQLESKREDDQREHS 124  
 DB 380 SSPQMAQHNNMVDGSGKVEIWRVENN 406  
 RESULT 29  
 ID DNL1 CABEL STANDARD; PRT; 847 AA.  
 AC 027474;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA ligase (EC 6.5.1.1) (Polydeoxyribonucleotide synthase [ATP]).  
 GN Lig-1 OR C29A12.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OK NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Wiltonson J.:  
 RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.  
 RL FUNCTION: THIS PROTEIN SEALS DURING DNA REPLICATION, DNA  
 RL RECOMBINATION AND DNA REPAIR NICKS IN DOUBLE-STRANDED DNA.  
 CC -1- CATALYTIC ACTIVITY: ATP + (deoxyribonucleotide) (N) +  
 CC {deoxyribonucleotide} (N) = AMP + diphosphate +  
 CC {deoxyribonucleotide} (N+M).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: Z73970; CAA98242.1; -;  
 DR PIR: T19544; T19544.  
 DR WormPep: C29A12.3; CE05328.  
 DR InterPro: IPR00977; DNA\_ligase.  
 DR Pfam: PF01068; DNA\_ligase; 1.  
 DR Pfam: PF04679; DNA\_ligase\_A\_C; 1.  
 DR Pfam: PF04675; DNA\_ligase\_A\_N; 1.  
 DR TIGRfam: TIGR00574; dnl1.1.  
 DR PROSITE: PS00697; DNA\_LIGASE\_A1; 1.  
 DR PROSITE: PS00333; DNA\_LIGASE\_A2; 1.  
 DR PROSITE: PS0160; DNA\_LIGASE\_A3; 1.  
 DR DNA repair; DNA replication; DNA recombination; Cell division; Ligase;  
 KM ATP-binding. 363 363  
 FT BINDING 363 363  
 FT AMP (BY SIMILARITY).



SEQUENCE 847 AA; 94501 MW; A32B7D279F91002F CRC64;  
Query Match 10.4%; Score 83; DB 1; Length 847;  
Best Local Similarity 26.2%; Pred. No. 56;  
Matches 44; Conservative 26; Mismatches 68; Indels 30; Gaps 8;  
QY 1 DPGVESELKPHRVTVTITONGKEMSTIVSEDFILPYKGE-----LEKGYQFGWEI 53  
DB 656 DDCGKISLPRPR-LIRDPKNSDDATSSQ--VLEMYKQGFANQKIEKADAVD--ED 710  
QY 54 SGFEKGKADGYVNL-----SKDTFIKPVFKIEKKEENKPTFDVSKK--DNPOV 104  
DB 711 DEFEKEDEBEELMTNVSQSGSKENPVK-----EIKKETPK---SVSPKFEKKPPV 761  
QY 105 NNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNISSKSTNNP 152  
DB 762 KSPVVKSPVKSPPKIKKEAKKKGPVASIFSSSTKKNEDKVKESDP 809  
RESULT 30  
GLNA\_PYRAB STANDARD; PRT; 439 AA.  
ID GLNA\_PYRAB  
AC Q9UY99;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
GN GLNA OR PYRAB1090 OR PAB1292.  
OS Pyrococcus abyssi.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
RN NCBI\_TaxID=29292;  
RX STRAIN=FROM N.A.  
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,  
Roch O., Prieur D., Querellon J., Ripp R., Thierry J.-C.,  
Van der Oost J., Weissendach J., Zivanovic Y., Forterre P.;  
RT "An integrated analysis of the genome of the hyperthermophilic  
archaeon Pyrococcus abyssi.";  
RL Mol. Microbiol. 47:1495-1512(2003).  
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
L-glutamine.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AJ248288; CAB50513.1; -.  
DR PIR; C75009; C75009.  
DR HSSP; P06201; 1LGR.  
DR InterPro; IPR001691; GLN synth.  
DR InterPro; IPR004809; GLNA.  
DR InterPro; IPR001637; GLNA adenyltn.  
DR Pfam; PF00120; gln-synt; 1.  
DR Pfam; PF03951; gln-synt N; 1.  
DR ProDom; PD001057; Gln\_synt\_C; 1.  
DR TIGRfam; TIGR00653; GLNA; 1.  
DR PROSITE; PS00180; GLNA\_1; 1.  
DR PROSITE; PS00181; GLNA\_ATP; 1.  
KW ligase; Complete proteome.  
FT BINDING 358 AMP (UNDER CONDITIONS OF ABUNDANT  
GLUTAMINE) (BY SIMILARITY).  
SQ SEQUENCE 439 AA; 49742 MW; 62CCFD3970A98AF0 CRC64;  
Query Match 10.3%; Score 82.5; DB 1; Length 439;

Best Local Similarity 17.8%; Pred. No. 30;  
Matches 28; Conservative 24; Mismatches 36; Indels 69; Gaps 4;  
QY 42 LEKGYQFGDWEISGEGKKDAGVNLISDTFI-----KP----- 76  
DB 42 IEDGISFDDSSVPGQIEDSDILFKADPDITVEVPMNVAHVGYIKDGRPGADPRG 101  
QY 77 VFKEIEKKEEN-----KPTFDVSKKDN----- 101  
DB 102 VLKRVTEKLAENGIRAYIGPEPEFYLFKNGSWELEIPVGGYFDLITDKAKDKREIA 161  
QY 102 -----POVNSQLNESHKEDLQREHSQKSDS 129  
DB 162 EYMPSEGLVPEVLHHEVGAQHEIDFRYDEALKTADN 198  
RESULT 31  
HS70\_PYRSA STANDARD; PRT; 649 AA.  
ID HS70\_PYRSA  
AC P37899;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 01-OCT-1994 (Rel. 30, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Heat shock 70 kDa protein.  
GN HSP70.  
OS Pyrenomonas salina.  
OC Nucleomona.  
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Pyrenomonas.  
OC NCBI\_TaxID=3034;  
RN (1)  
RX SEQUENCE FROM N.A.  
RA MEDLINE=94268506; PubMed=8208251;  
RA Hofmann C.J.B., Rensing S.A., Hauber M.M., Martin W.F., Mueller S.B.,  
RA Couch J., McFadden G.I., Igloi G., Water U.-G.,  
RT "The smallest known eukaryotic genomes encode a protein gene: towards  
an understanding of nucleomorph functions.";  
RL Mol. Genet. 243:600-604(1994).  
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X72621; CAA51197.1; -.  
DR PIR; S42488; S42488.  
DR HSSP; P08109; 1CKR.  
DR InterPro; IPR001023; Hsp70.  
DR Pfam; PF00012; HSP70; 1.  
DR PRINTS; PR00301; HEATSHOCK70.  
DR ProDom; PD000089; HSP70; 1.  
DR PROSITE; PS00297; HSP70\_1; 1.  
DR PROSITE; PS00329; HSP70\_2; 1.  
DR PROSITE; PS01036; HSP70\_3; 1.  
DR ATP-binding; Heat shock; Nucleomorph.  
SQ SEQUENCE 649 AA; 72079 MW; B627B08F90C9164 CRC64;  
Query Match 10.3%; Score 82.5; DB 1; Length 649;  
Best Local Similarity 23.8%; Pred. No. 46;  
Matches 34; Conservative 25; Mismatches 43; Indels 41; Gaps 7;  
QY 9 KPHRVTVTITONG-----KEMSTIVSEDFILPYKGELEKGYQFGDWEISGEGKKDA--- 62  
DB 503 KSNKTIITNDKRLSGSEIERVVEAE-----KYKTEDEK-----LKKLEAKNS 547  
QY 63 --GYVINS---KDTFIKPVFKIEKKEENKPTFDVSKKDNPOVNSQLNESHKED 117  
DB 548 LENYAVANINHWVD-----EKLKEKIQEEDKKSIIEKYKE-----VLEFIETNED 592  
QY 118 LQREHSQKSDSTKDVATVLDK 140



Db 593 LEKEVEKEKEKRFANPIISK 615

RESULT 32  
MAPB\_RAT STANDARD; PRT; 2459 AA.  
AC P15305; 062958; Q9ER21; Q9QW92;  
BT 01-APR-1990 (Rel. 14, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Microtubule-associated protein 1B (MAP1B) [Contains: MAP1  
light chain LC1].  
GN MAP1B.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;  
RN [1]  
RP SEQUENCE OF 1-142 FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Testis;  
RX MEDLINE=96257242; PubMed=8666295;  
RA Liu D., Fischer I.;  
RT "Isolation and sequencing of the 5' end of the rat microtubule-  
associated protein (MAP1B)-encoding cDNA.";  
RL Gene 172:307-308(1996).  
RN [2]  
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.  
RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Gli1 tumor;  
RX MEDLINE=92347374; PubMed=1639092;  
RA Zanner W., Kratz J., Staunton J., Feick P., Wiche G.;  
RT "Identification of two distinct microtubule binding domains on  
recombinant rat MAP1B.";  
RL Eur. J. Cell Biol. 57:66-74(1992).  
RN [3]  
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.  
RC TISSUE=Spinal cord;  
RX MEDLINE=90059671; PubMed=2555150;  
RA Rientitz A., Gremmling G., Hermann-Borgmeyer I., Kirsch J.,  
RA Litzner U.Z., Pflor P., Gundelfinger E.D., Schmitt B., Betz H.;  
RT "Neuraxin, a novel putative structural protein of the rat central  
nervous system that is immunologically related to microtubule-  
associated protein 5.";  
RL EMBO J. 8:2879-2888(1989).  
RN [4]  
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.  
RX MEDLINE=97405699; PubMed=9260743;  
RA Ma D., Nothias F., Boyne L.J., Fischer I.;  
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)  
in rat CNS and PNS during development.";  
RL J. Neurosci. Res. 49:319-332(1997).  
RN [5]  
RP FUNCTION: The function of brain MAPs is essentially unknown.  
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes  
that accompany neurite extension. Possibly MAP1B binds to at least  
two tubulin subunits in the polymer, and this bridging of subunits  
might be involved in nucleating microtubule polymerization and in  
stabilizing microtubules.  
CC SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate  
with MAP1A and MAP1B proteins.  
CC TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,  
cerebellum and cerebrum). Not expressed in liver, spleen, kidney,  
heart or muscle.  
CC DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic  
nerve levels are high early in development but decrease during  
postnatal development and are low in adults. In dorsal root  
ganglia levels remain high throughout development.  
CC INDUCTION: By nerve growth factor.  
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence  
KKKK and KKKI/V, repeated but not at fixed intervals, which is  
responsible for the binding of MAP1B to microtubules.  
CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated  
from MAP1B by proteolytic processing. It is free to associate with  
both MAP1A and MAP1B. It interacts with the amino-terminal region

CC of MAP1B (By similarity).  
CC -1- PTM: Phosphorylated.  
CC -1- SIMILARITY: TO MAP1A.  
CC -1- CAUTION: A C-terminal fragment of this protein (residues 1597 to  
2459) was originally described as neuraxin in Ref.3.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb.ch](mailto:license@isb.ch)).  
CC -----  
DR EMBL: U52950; AAB17068.1; -;  
DR EMBL: X60370; CAC16162.1; -;  
DR EMBL: X16623; CAA34620.1; ALT\_SEQ.  
DR PIR: A56577; A56577.  
DR InterPro: IPR000102; MAP1B\_neuraxin.  
DR Pfam: PF00414; MAP1B\_neuraxin: 10.  
DR PROSITE: P800230; MAP1B\_NEURAXIN; 8.  
DR Microtubules; Repeat; Phosphorylation.  
KW CHAIN ? 2459 MAP1 LIGHT CHAIN LC1.  
FT REPEAT 1869 1885 MAP1B 1.  
FT REPEAT 1886 1902 MAP1B 2.  
FT REPEAT 1903 1919 MAP1B 3.  
FT REPEAT 1920 1936 MAP1B 4.  
FT REPEAT 1937 1953 MAP1B 5.  
FT REPEAT 1954 1970 MAP1B 6.  
FT REPEAT 1988 2004 MAP1B 7.  
FT REPEAT 2005 2021 MAP1B 8.  
FT REPEAT 2022 2038 MAP1B 9.  
FT REPEAT 2039 2055 MAP1B 10.  
FT DOMAIN 559 1035 GLU-RICH.  
FT DOMAIN 588 786 LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
LYS-RICH.  
FT DOMAIN 2224 2312 LYS-RICH.  
FT CONFLICT 127 127 M -> V (IN REF. 1).  
FT CONFLICT 140 140 T -> S (IN REF. 1).  
FT CONFLICT 2112 2112 R -> K (IN REF. 3).  
FT CONFLICT 2169 2169 L -> I (IN REF. 3).  
SQ SEQUENCE 2459 AA; 269497 MW; 2E3F6872DEDB8BA2 CRC64;  
Query Match 10.3%; Score 82.5; DB 1; Length 2459;  
Best Local Similarity 25.4%; Pred. No. 1.9e+02;  
Matches 35; Conservative 25; Mismatches 49; Indels 29; Gaps 6;  
QY 20 GENESTIVSEDFILPYVKGLEKGYQFDGWEISGFEGKTDAGVYINLSKDTFIKPVFK 79  
DB 593 GKVESKPSVTEKE-----VPSKEEGSPVAAVAEKAKATESK-----KVTXDKVYKKEIK 642  
QY 80 -KIEKKEEENKPTFDVSKKQNDPVNHSQINSHRKEDLOREHSQKSDSTQVATVVL 138  
DB 643 TKPEEKKEE--KPKKEVAKED-----KTPLKDEKPKKEAKKEIKKEIK 686  
QY 139 --DKNNISSKSTNNPNK 154  
DB 687 KEKKEKKEKVEKKEKTPK 704  
RESULT 33  
YMB4\_YEAST STANDARD; PRT; 279 AA.  
AC P49957;  
BT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Hypothetical 32.4 kDa protein in TAP40-ERV25 intergenic region.  
GN YMD14W OR YM9571.04.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxId=4932;

```

RN SEQUENCE FROM N.A.
RP STRAIN=6288C / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jager K., Lye G., Mole S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrall B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII."
RL Nature 387:90-93(1997).
CC -1- SIMILARITY: TO S.POMBE SPAC13D6.03C AND SOME, TO C.ELEGANS
CC C14B1.5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z49810; CAA8938.1; -.
CC PIR; S55105; S55105.
CC SGD; S0004476; YML014W.
CC InterPro; IPR001601; Methyltransf.
CC InterPro; IPR000051; SAM_bind.
CC Hypothetical protein.
CC SEQUENCE 279 AA; 32438 MW; 2F8795BB3C640D3B CRC64;
SQ
Query Match 10.3%; Score 82; DB 1; Length 279;
Best Local Similarity 28.9%; Pred. No. 20;
Matches 24; Conservative 14; Mismatches 39; Indels 6; Gaps 3;
QY 51 WEISGEKKDAGYVNLSDRTFKYFKIEKKEENKPTDVSKKKDPVNHSQLN 110
DB 145 WALE-QGSSRRGYHEGMEDVFPWVLPKSKSKPKTKSP--AKVTRPKNLMNIP 199
QY 111 ESHRKEDLOR-EHHSQKSDSTKD 132
DB 200 PKERSEYLORKEEOQRSKSLDD 222
ID YKTS_YEAST STANDARD; PRT; 427 AA.
AC P36046;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Hypothetical 47.4 kDa protein in PAB1-MST1 intergenic region.
GN YKL195W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z28195; CAA82039.1; -.
CC PIR; S38032; S38032.
CC SGD; S0001678; YKL195W.

```

```

KM Hypothetical protein.
SQ SEQUENCE 427 AA; 47416 MW; 44386D250DE5DED4 CRC64;
Query Match 10.3%; Score 82; DB 1; Length 427;
Best Local Similarity 27.4%; Pred. No. 35;
Matches 29; Conservative 18; Mismatches 43; Indels 16; Gaps 3;
QY 41 ELEKGYQFGWIEGEGKKDAGYVNLSDRTFKYFKIEKKEENKPTDVSKKKD 100
DB 142 ETEAGQGLGDDXIGASKVADEGLVLAEDN-----KSSPDKOTDESK-----VSTKDD 191
QY 101 NPQVNHSQLNESHREKEDIQREHHSQKSDSTKDVATVLDKNNISRK 146
DB 192 -----EQSNEDNATANNOKDENISENSEENTSKTLDDNAGSSSE 231
ID GLNA_PYRFU STANDARD; PRT; 439 AA.
AC 005907;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
GN GLNA OR PF0450.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RX MEDLINE=94365840; PubMed=7916055;
RA Brown J.R., Masuchi Y., Robb F.T., Doolittle W.F.;
RT "Evolutionary relationships of bacterial and archaeal glutamine
RT synthetase genes."
RL J. Mol. Evol. 38:566-576(1994).
ID [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RX Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome."
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
CC L-glutamine.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L12410; AAJ1968.1; -.
CC EMBL; AE010168; AAL80574.1; -.
CC HSSP; P06201; ILGR.
CC InterPro; IPR001691; GLN_synth.
CC InterPro; IPR004809; GLNA.
CC InterPro; IPR001637; GLNA_adenyltn.
CC Pfam; PF00120; gln-synt_1.
CC Pfam; PF03951; gln-synt_N; 1.
CC ProDom; PD001057; Gln_synt_C; 1.
CC TIGRPFAMs; TIGR00653; GLNA; 1.
CC PROSITE; PS00180; GLNA_1; 1.
CC PROSITE; PS00181; GLNA_ATP; 1.
CC Ligase; Complete proteome.
CC BINDING 358
FT FT
FT BINDING 203 203 AMP (UNDER CONDITIONS OF ABUNDANT
FT CONFLICT 232 232 GLUTAMINE) (BY SIMILARITY).
FT K -> G (IN REF. 1).
FT M -> L (IN REF. 1).

```

FT CONFLICT 351 351 A -> S (IN REF. 1).  
 FT CONFLICT 370 370 I -> L (IN REF. 1).  
 FT CONFLICT 376 376 S -> N (IN REF. 1).  
 FT CONFLICT 382 382 E -> G (IN REF. 1).  
 FT CONFLICT 397 397 A -> S (IN REF. 1).  
 FT CONFLICT 421 426 MPDTR -> IPPDTE (IN REF. 1).  
 SQ SEQUENCE 439 AA; 50183 MW; 96563ABE9A0E0892 CRC64;

Query Match 10.3%; Score 82; DB 1; Length 439;  
 Best Local Similarity 18.1%; Pred. No. 33;  
 Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

QY 45 GYQFGMEISGPEGKKDAGYVNLKSKTEI-----KP-----VFK 79  
 DB 45 GISFGSSVPGQIGEDSLVFKADPTVEVPMNVARVYGYKDKNPGADPGILK 104  
 QY 80 KIEEKEEN-----KPTFVSKKKN----- 101  
 DB 105 RALEEKEGKAYVIGPEPEFLFKNGTWELIPDVGVFDILTLDKARDIRREIAEYM 164  
 QY 102 -----PQVNSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 143  
 DB 165 PSFGILPEVLHHEVGKAOHEIDFRYDEALKTADNIVSFYITKAVAEHMG 215

## RESULT 36

GLNA\_PYRO STANDARD; PRT; 439 AA.

AC P36687;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glutamine synthetase I (EC 6.3.1.2) (Glutamate--ammonia ligase) (GSI).  
 GN GLNA.

OS Pyrococcus woesei.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.

OK NCBI\_TaxID=2262;

RN [1]

RP SEQUENCE FROM N.A.  
 RA MEDLINE=93259940; PubMed=8098326;  
 RA Tiboni O., Cammarano P., Sanangelantoni A.M.;

RT "Cloning and sequencing of the gene encoding glutamine synthetase I  
 from the archaeum Pyrococcus woesei: anomalous phylogenies inferred  
 from analysis of archaeal and bacterial glutamine synthetase I  
 sequences.";

RT J. Bacteriol. 175:2961-2969 (1993).

CC -1- CARBLYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
 L-glutamine.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, X60161; CAA42730.1; -

DR PIR; A36911; A36911.

DR HSP; P06201; ILGR.

DR InterPro; IPR001691; GLN\_synth.

DR InterPro; IPR004809; GLN.

DR InterPro; IPR001637; GLN\_adenyltn.

DR Pfam; PF00120; gln-synt; 1.

DR Pfam; PF03951; gln-synt\_N; 1.

DR ProDom; PD001057; Gln\_synt\_C; 1.

DR TIGRFAMs; TIGR00653; GLNA.1.

DR PROSITE; PS00180; GLNA.1; 1.

DR PROSITE; PS00181; GLNA\_ATP; 1.  
 KM ligase.

FT BINDING 358 358 AMP (UNDER CONDITIONS OF ABUNDANT  
 FT SEQUENCE 439 AA; 50066 MW; C503BA3B790BD25C CRC64;  
 SQ SEQUENCE

Query Match 10.3%; Score 82; DB 1; Length 439;  
 Best Local Similarity 18.1%; Pred. No. 33;  
 Matches 31; Conservative 26; Mismatches 42; Indels 72; Gaps 5;

QY 45 GYQFGMEISGPEGKKDAGYVNLKSKTEI-----KP-----VFK 79  
 DB 45 GISFGSSVPGQIGEDSLVFKADPTVEVPMNVARVYGYKDKNPGADPGILK 104  
 QY 80 KIEEKEEN-----KPTFVSKKKN----- 101  
 DB 105 RALEEKEGKAYVIGPEPEFLFKNGTWELIPDVGVFDILTLDKARDIRREIAEYM 164  
 QY 102 -----PQVNSQLNESHKEDLQREHSQKSD---STKDVATVLDKNNI 143  
 DB 165 PSFGILPEVLHHEVGKAOHEIDFRYDEALKTADNIVSFYITKAVAEHMG 215

## RESULT 37

CENC\_HUMAN STANDARD; PRT; 943 AA.

AC 003188; OSPOM5;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Centromere protein C (CENP-C) (Centromere autoantigen C).

GN CENPC1 OR CENPC.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RX MEDLINE=92323541; PubMed=1339310;

RA Satoh H., Tomkiel J., Cooke C.A., Rattie H. III, Maurer M.,

RA Rothfield N.F., Barnshaw W.C.;

RT "CENP-C, an autoantigen in scleroderma, is a component of the human

inner kinetochore plate.";

RT Cell 70:115-125 (1992).

RN [2]

RP SEQUENCE OF 1-21 FROM N.A.

RA Poppe M., Boltz J., Hahn B., Dobat K., Eickelbaum W., Paweletz N.,

RA Arand M., Knehr M.;

RT "Promoter characterization of centromere protein C reveals its

participation in cell cycle regulation in late G1-phase and expression

control by E2F-1, pRB, p107 and Sp-1.";

RT Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: Component of the inner kinetochore plate. Required for

normal kinetochore assembly.

CC -1- SUBUNIT: Binds to DAXX.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, M95724; AAF51974.1; -

DR EMBL; AF151723; AAF73191.1; -

DR PIR; A42681; A42681.

DR Genew; HGNC:1854; CENPC1.

DR MIM; 117141; -

DR GO; GO:0005699; C:kinetochore; TAS.

DR GO; GO:0003677; F:DNA binding activity; TAS.

DR InterPro; IPR007113; Cupin sup.

DR Chromosomal protein; Nuclear protein; DNA-binding; Phosphorylation;

KM Centromere. 259 273 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 484 499 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 558 574 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 780 798 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT MOD RES 75 75 PHOSPHORYLATION (POTENTIAL).  
 FT MOD RES 732 732 PHOSPHORYLATION (POTENTIAL).  
 SQ SEQUENCE 943 AA; 106925 MW; 6D80810A3E476376 CRC64;

Query Match 10.3%; Score 82; DB 1; Length 943;  
 Best Local Similarity 23.2%; Pred. No. 75;  
 Matches 44; Conservative 18; Mismatches 76; Indels 52; Gaps 7;

QY 3 GEVSEIKPRHV-IVTQNGKEMSSSTIVSEDFILPYKGLKGYFDGMEISGFGKDD 61  
 DB 343 GRKSRREHNI LKPTLANDHSHKHPVETS--QPSDKTVLDTSYALIDETVNNVSTKY 400  
 QY 62 AGVIVNLSKDTFKIPVKFKEKK-----EENKPTFDVSKKD-NPQVNSQLNESHR 114  
 DB 401 EWSKNAEKRSKRRTIKQKQKRFMAKPAEQ-----LDVGQKDNINHTSHITQDEPQR 456  
 QY 115 KEDLQREH-----SQKSDSTKD-----VTAT 136  
 DB 457 NSDRNMEHEHMGNDGVSKQMPVGSKKSTRKDESKKRFSSSKNKLVEEVTST 516  
 QY 137 VLDKNNISSK 146  
 DB 517 VTKSRRISSR 526

RESULT 38  
 IP2\_HELPJ STANDARD; PRT; 949 AA.  
 ID IP2\_HELPJ  
 AC Q9ZM46;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Translation initiation factor IF-2.  
 GN INFB OR JHP0377.  
 OS Helicobacter pylori J99 (Campylobacter pylori J99).  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 OC Helicobacteraceae; Helicobacter.  
 OX NCBI\_TaxID=85963;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=99120557; PubMed=9923682;  
 RA Alm R.A., Ling L.-S.L., Molt D.T., King B.L., Brown E.D., Doig P.C.,  
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,  
 RA Thummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,  
 RA Gibson R., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,  
 RA Trust T.J.;  
 RT Genomic sequence comparison of two unrelated isolates of the human  
 RT gastric pathogen Helicobacter pylori.";  
 RL Nature 397:176-180(1999).  
 CC -1- FUNCTION: One of the essential components for the initiation of  
 CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous  
 CC hydrolysis and promotes its binding to the 30S ribosomal subunits.  
 CC Also involved in the hydrolysis of GTP during the formation of the  
 CC 70S ribosomal complex (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE001472; AAD05948.1; -;  
 DR PIR; E71940; E71940.  
 DR HAMAP; MF\_00100; -; 1.

DR Interpro; IPR000795; EF\_GTPbind.  
 DR Interpro; IPR004161; EFTU\_D2.  
 DR Interpro; IPR000178; IF2\_D.  
 DR Interpro; IPR006847; IF2\_N.  
 DR Interpro; IPR005225; Small\_GTP.  
 DR Pfam; PFO0009; GTP\_EFTU; 1.  
 DR Pfam; PFO3144; GTP\_EFTU\_D2; 2.  
 DR Pfam; PFO4760; IF2\_N; 1.  
 DR ProDom; PD186100; IF2; 1.  
 DR TIGRfam; TIGR00487; IF-2; 1.  
 DR TIGRfam; TIGR00231; small\_GTP; 1.  
 DR PROSITE; PS01176; IF2; 1.  
 KW Initiation factor; Protein biosynthesis; GTP-binding;

KM Complete proteome.  
 FT DOMAIN 451 599 G-DOMAIN.  
 FT NP\_BIND 457 464 GTP (BY SIMILARITY).  
 FT NP\_BIND 503 507 GTP (BY SIMILARITY).  
 FT NP\_BIND 557 560 GTP (BY SIMILARITY).  
 FT DOMAIN 321 326 POLY-GLU.  
 SQ SEQUENCE 949 AA; 105961 MW; FA8969B0C64B3278 CRC64;

Query Match 10.3%; Score 82; DB 1; Length 949;  
 Best Local Similarity 28.7%; Pred. No. 75;  
 Matches 29; Conservative 20; Mismatches 20; Indels 32; Gaps 5;

QY 70 KDTFKIPVKFKIEKKKEENKPTFDVSKKD-----NPO-VNHSQLES 112  
 DB 98 EETKQPKPKTKREKKKEAPAP--IIRKKEIYNTFENQTPVNTPKAVSHQIERA 154  
 QY 113 HRKEDLQREHSSQKSDSTQVATVLDKNNISSKSTNNPN 153  
 DB 155 -KQKLGELQKREAA-----LNKLTQSNVTNTNAN 183

RESULT 39  
 RBP2\_PLAVB STANDARD; PRT; 1251 AA.  
 ID RBP2\_PLAVB  
 AC Q00759;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Reticulocyte binding protein 2 (Fragment).  
 GN RBP2.  
 OS Plasmodium vivax (strain Belem).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=31273;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=92315338; PubMed=1617731;  
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
 RA "A reticulocyte-binding protein complex of Plasmodium vivax  
 RT merozoites.";  
 RT Cell 69:1213-1226(1992).  
 RL Cell 69:1213-1226(1992).  
 CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO  
 CC HUMAN RETICULOCYTE CELLS.  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M88098; AAA29744.1; -;  
 DR Malaria; Receptor; Membrane.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 FT NON\_TER 1251  
 SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 10.3%; Score 82; DB 1; Length 1251;  
 Best Local Similarity 22.9%; Pred. No. 1e+02;

Matches	39; Conservative	35; Mismatches	66; Indels	30; Gaps	8;
Qy	5 VSELKPH-RATYVTIQNGKMSSTI-----VSESDITLPYKGELEKQYQDWEISG	55			
Db	266 LSEIKKEYDKDCTTEISNSKRGKDIPEFEKFKPENSNSNKNINIEINIRNSBOYLD	325			
Qy	56 FE-QKGDGVYINL--SKDTFKIPVFKK----IEKKKEENKPFYDYSK--KKONPOV-	104			
Db	326 IEDAEKASTVVELFHKEHTTISNIFKSESLGVETKCKKINAKEDIMKEIERINSIQ	385			
Qy	105 -----NHSOLNESHRRKEDLOREHSHQSDSTDYATATVD--KNNS 144				
Db	386 TGVGFQGNLAKNKNPFHYDNAEDDLNDKSNNAVVLLETNLESYKHLIS 435				
RESULT 40					
CRT1_ARATH					
ID CRT1_ARATH	STANDARD;	PRT;	425 AA.		
AC	004151.				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Calreticulin 1 precursor.				
GN	CRT1 OR ARI656340 OR F13N6.20 OR F14G9.5.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosides II; Brassicales; Brassicaceae; Arabidopsie.				
OX	NCBI_Taxid=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97303616; PubMed=9159940;				
RA	Nelson D.E., Glaunsinger B., Bohert H.J.;				
RT	"Abundant accumulation of the calcium-binding molecular chaperone				
RT	calreticulin in specific floral tissues of Arabidopsis thaliana."				
RL	Plant Physiol. 114:29-37(1997).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Columbia;				
RX	MEDLINE=21016719; PubMed=1130712;				
RA	Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,				
RA	White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,				
RA	Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,				
RA	Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,				
RA	Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,				
RA	Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizlar L.,				
RA	Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,				
RA	Kim C.-H., Koo H.D., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,				
RA	Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.-P.,				
RA	Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maiti R., Marziani A.,				
RA	Miltschker J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,				
RA	Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,				
RA	Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,				
RA	Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,				
RA	Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,				
RA	Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;				
RT	"Sequence and analysis of chromosome 1 of the plant Arabidopsis				
RT	thaliana."				
RL	Nature 408:816-820(2000).				
CC	-1- FUNCTION: THIS PROTEIN BINDS CALCITUM. THERE ARE BOTH HIGH AND				
CC	LOW AFFINITY CALCITUM-BINDING SITES (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).				
CC	-1- SIMILARITY: BELONGS TO THE CALRETICULIN FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
CC	EMBL; U66343; AAC49695.1; -				

	DR	EHBL; AC058785; AAG51504.1; -.	
	DR	EMBL; AC069159; AAG50908.1; -.	
	DR	PfR; C96605; C96605.	
	DR	SWISS-2DPAGE; O04151; ARATH.	
	DR	InterPro; IPR001580; Calreticulin.	
	DR	InterPro; IPR000886; ER_target.	
	DR	Pfam; PF00262; calreticulin.1.	
	DR	PRINTS; PR00626; CALRETICULIN.	
	DR	ProDom; PD001866; Calreticulin.1.	
	DR	PROSITE; PS00014; ER_TARGET.1.	
	DR	PROSITE; PS00803; CALRETICULIN.1.	
	DR	PROSITE; PS00804; CALRETICULIN.2; 1.	
	DR	PROSITE; PS00805; CALRETICULIN REPEAT; 2.	
	KW	Endoplasmic reticulum; Calcium-binding; Repeat; Signal; Glycoprotein;	
	KW	Multigene family.	
FT	SIGNAL	1 22	POTENTIAL.
FT	CHAIN	23 425	CALRETICULIN 1.
FT	CARBOHYD	59 59	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	154 154	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	399 399	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	422 425	PREVENT SECRETION FROM ER (POTENTIAL).
SQ	SEQUENCE	425 AA; 48527 MW; BCC0BEE2342642E CRC64;	.
Qy	Query Match	10.2%; Score 81.5; DB 1; Length 425;	
Dy	Best local Similarity	21.6%; Pred. No. 35;	
Bd	Matches 32; Conservative	28; Mismatches 63; Indels 25; Gaps 5	
Oy	7 ELKEHHVYTTIQNGKMSSTIVSEED-----FILPYRK-----GELKGYPQRDMGEIS	54	
Dy	278 EWPKPTKRIPAYKSGAKAMIDNPEFKDDDELVIYFPLKIYGVGLMOWKSGSLFDNVLVLS	337	
Oy	55 GFEGKKADAGVINLSDTF-----IKPVFKIEEKKEENKPTFDVSKKKDNPQYNHSQ	108	
Dy	338 -----DDPEYAKKLAEFTWGKHKADEAKAFADEAKCKKEEBESK--DAPAESDAEBAEDD	390	
Oy	109 LNSSHKEDLRREHSSQSXSDTKYVTAT	136	
Dy	391 DNEGDDSDNESKSSEETKEAEETKEEET	418	

Search completed: February 10, 2004, 10:49:53  
Job time : 8:15925 secs

DR EMBL; U66343; AAC49695.1; -

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 27.2661 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799  
Sequence: 1 DTGEVSELEKPRVTVTTONG.....ATVLDKNKISSKSTNNPNK 154

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL.23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriaph:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	799	100.0	2119	2 Q9AHT5	Q9AHT5 streptococc
2	799	100.0	2140	16 Q97RY6	Q97RY6 streptococc
3	796	99.6	2144	16 Q8DDP7	Q8DDP7 streptococc
4	795	99.5	2144	2 Q984M8	Q984M8 streptococc
5	118	14.8	775	16 Q8CPK8	Q8CPK8 streptococc
6	112.5	14.1	361	5 Q95P15	Q95P15 plasmodium
7	112.5	14.1	379	5 Q9U6C4	Q9U6C4 plasmodium
8	111.5	14.0	379	5 Q25706	Q25706 plasmodium
9	110.5	13.8	379	5 Q25705	Q25705 plasmodium
10	110	13.8	346	5 Q9U0G0	Q9U0G0 plasmodium
11	109	13.6	380	5 Q26019	Q26019 plasmodium
12	109	13.6	3008	5 Q81436	Q81436 plasmodium
13	108.5	13.6	600	5 Q77355	Q77355 plasmodium
14	108.5	13.6	1038	13 Q90784	Q90784 gallus galli
15	108	13.5	354	5 Q25995	Q25995 plasmodium
16	108	13.5	354	5 Q81J55	Q81J55 plasmodium

17	107.5	13.5	470	10 Q9FJK9	Q9FJK9 arabidopsis
18	107	13.4	829	5 Q815F3	Q815F3 plasmodium
19	105	13.1	951	5 Q96229	Q96229 plasmodium
20	103.5	13.0	325	5 Q44016	Q44016 dictyosteli
21	102.5	12.8	2081	10 Q9LH98	Q9LH98 arabidopsis
22	101.5	12.7	382	5 Q9V770	Q9V770 drosophila
23	101.5	12.7	556	5 Q9V779	Q9V779 drosophila
24	101.5	12.7	785	5 Q9G082	Q9G082 drosophila
25	101	12.6	329	5 Q9NFW9	Q9NFW9 plasmodium
26	100	12.5	312	16 Q9PPL5	Q9PPL5 campylobact
27	100	12.5	375	4 Q14712	Q14712 homo sapien
28	100	12.5	1130	5 Q81JZ4	Q81JZ4 plasmodium
29	99	12.4	211	5 P91488	P91488 caenorhabdi
30	99	12.4	1859	5 Q81C27	Q81C27 plasmodium
31	99	12.4	2563	5 Q813A0	Q813A0 plasmodium
32	98.5	12.3	238	5 Q812Z6	Q812Z6 plasmodium
33	98	12.3	157	5 Q9VOV0	Q9VOV0 drosophila
34	98	12.3	4524	5 Q81J39	Q81J39 plasmodium
35	97.5	12.2	449	5 Q81HW3	Q81HW3 plasmodium
36	97.5	12.2	556	5 Q95S93	Q95S93 drosophila
37	97.5	12.2	614	16 Q9K5S1	Q9K5S1 bacillus ha
38	97	12.1	385	5 Q93424	Q93424 caenorhabdi
39	97	12.1	558	5 Q815T1	Q815T1 plasmodium
40	97	12.1	988	10 Q9STN4	Q9STN4 arabidopsis
41	97	12.1	3127	5 Q81DA0	Q81DA0 plasmodium
42	96.5	12.1	540	10 Q94C59	Q94C59 arabidopsis
43	96.5	12.1	540	10 Q9S484	Q9S484 arabidopsis
44	96	12.0	421	16 Q8E0Q7	Q8E0Q7 shewanella
45	96	12.0	531	5 Q81426	Q81426 plasmodium

## ALIGNMENTS

## RESULT 1

Q9AHT5 PRELIMINARY: PRT; 2119 AA.

AC Q9AHT5 01-JUN-2001 (TREMBLrel. 17, Created)  
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Serine protease (Fragment).  
GN PRTA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
RA Witzemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,  
RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,  
RA Gayle A., Brewah Y.A., Waleh W., Barren P., Lathigra R., Hanson M.,  
RA Langermann S., Johnson S., Koenig S.;  
RT "Use of a whole Genome Approach to Identify Vaccine Molecules  
Affording Protection against Streptococcus pneumoniae Infection.";  
RT Infect. Immun. 69:1593-1598(2001).  
RL -I- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
CC AN AMIDE BOND (BY SIMILARITY).  
CC EMBL: AF291699; AAK19159.1; -  
DR HSSP; P00782; 2SRT.  
DR MEROPS; S08.064; -  
DR INTERPRO; IPR001899; Gram\_pos\_anchor.  
DR INTERPRO; IPR006192; LPXTG.  
DR INTERPRO; IPR003137; PA.  
DR INTERPRO; IPR000209; Peptidase\_S8.  
DR INTERPRO; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRPFAMs; TIGR01167; LPXTG\_anchor; 1.



DR PROSITE; PS09847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
FT Cell wall; Peptidoglycan-anchor; Protease.  
NON TER 1  
SQ SEQUENCE 2119 AA; 238226 MW; 517F9B7F6B960A6A CRC64;  
Query Match 100.0%; Score 799; DB 2; Length 2119;  
Best Local Similarity 100.0%; Pred. No. 2.1e-51;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTGEVSELKPHRVYTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFGWELISGFEK 60  
DB 1932 DTGEVSELKPHRVYTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFGWELISGFEK 1991  
QY 61 DAGVYINLSKDTPIKVPFKKIEKKEENKPTFDVSKKKDNPQVNHSQLNESHKEDLOR 120  
DB 1992 DAGVYINLSKDTPIKVPFKKIEKKEENKPTFDVSKKKDNPQVNHSQLNESHKEDLOR 2051  
QY 121 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 154  
DB 2052 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 2085  
RESULT 2  
Q97RY6 PRELIMINARY; PRT; 2140 AA.  
ID Q97RY6;  
AC 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)  
DE Serine protease, subtilase family.  
GN SP0641.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RA SEQUENCE FROM N.A.  
RC STRAIN=TIGR4;  
RX MEDLINE=21357209; PubMed=11463916;  
RA Testelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
Holtaple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
R.A. Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
pneumoniae";  
RL Science 293:498-506(2001).  
DR EMBL; AE007373; AAK74791.1; -.  
DR MEROPS; S08\_064; -.  
DR TIGR; SP0641; -.  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR00209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00882; Peptidase\_S8; 2.  
DR PRINTS; PRO0723; SUBTILISIN.  
DR TIGRPFAM; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS09847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Protease; Complete proteome.

SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B314 CRC64;  
Query Match 100.0%; Score 799; DB 16; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 2.1e-51;  
Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTGEVSELKPHRVYTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFGWELISGFEK 60  
DB 1953 DTGEVSELKPHRVYTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFGWELISGFEK 2012  
QY 61 DAGVYINLSKDTPIKVPFKKIEKKEENKPTFDVSKKKDNPQVNHSQLNESHKEDLOR 120  
DB 2013 DAGVYINLSKDTPIKVPFKKIEKKEENKPTFDVSKKKDNPQVNHSQLNESHKEDLOR 2072  
QY 121 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 154  
DB 2073 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 2106  
RESULT 3  
Q8DOP7 PRELIMINARY; PRT; 2144 AA.  
ID Q8DOP7;  
AC Q8DOP7;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, last annotation update)  
DE Cell wall-associated serine proteinase Prta (Ec 3.4.21.-).  
GN PRta OR SP0561.  
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=171101;  
RN [1]  
RA SEQUENCE FROM N.A.  
RX MEDLINE=21429245; PubMed=11544234;  
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,  
Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,  
McIntire S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,  
Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,  
Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
Zook C.A., Baltz R.H., Jaskunas S.R., Roestek P.R. Jr., Skatrud P.L.,  
Glass J.I.;  
RT "Genome of the bacterium Streptococcus pneumoniae strain R6";  
RL J. Bacteriol. 183:5709-5717(2001).  
DR EMBL; AE008434; AAK9365.1; -.  
KW Hydrolyase; Complete proteome.  
SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4BD8C503A0C CRC64;  
Query Match 99.6%; Score 796; DB 16; Length 2144;  
Best Local Similarity 99.4%; Pred. No. 3.5e-51;  
Matches 153; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTGEVSELKPHRVYTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFGWELISGFEK 60  
DB 1957 DTGEVSELKPHRVYTIQNGKEMSSITVSEEDFILPVYKGELEKGYQFGWELISGFEK 2016  
QY 61 DAGVYINLSKDTPIKVPFKKIEKKEENKPTFDVSKKKDNPQVNHSQLNESHKEDLOR 120  
DB 2017 DAGVYINLSKDTPIKVPFKKIEKKEENKPTFDVSKKKDNPQVNHSQLNESHKEDLOR 2076  
QY 121 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 154  
DB 2077 EEHQSQSDSTKDVATATVLDKNNISSKSTNNPNK 2110  
RESULT 4  
Q9S4M8 PRELIMINARY; PRT; 2144 AA.  
ID Q9S4M8;  
AC Q9S4M8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)

DT 01-MAR-2003 (TRENBLREL. 23, last annotation update)  
DE Cell wall-associated serine proteinase Prta precursor.  
CN PRTA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OX Streptococcus.  
RN NCBI\_TaxId=1313;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3.B.  
RA Bette G., ten Thoren E., Bongaeerts R.J.M., Heinz H.-P., Zysk G.,  
RT "Cloning and sequencing of a novel surface protease of Streptococcus  
RT pneumoniae."  
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
CC AN AMIDE BOND (BY SIMILARITY).  
CC EMBL; AF127143; AAD4839.1; -.  
CC HSP; P00782; 2SPT.  
DR MEROPS; S08.064; -.  
DR InterPro; IPR001699; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPTTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR00209; Peptidase\_S8.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRFAMs; TIGR01167; LPTX anchor; 1.  
DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Cell wall; Peptidoglycan-anchor; Signal.  
FT SIGNAL 1  
FT CHAIN 20 2144 CELL WALL-ASSOCIATED SERINE PROTEINASE  
FT  
FT  
SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;  
PRTA.  
Query Match 99.5%; Score 795; DB 2; Length 2144;  
Best Local Similarity 98.7%; Pred. No. 4.2e-51;  
Matches 152; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DTGEVSELKPHRVVTYVTONGKEMSSITVSEEDFILPVYKGELEKGYQFGWEISGEK 60  
DB 1957 DTGEVSELKPHRITVTYVTONGKEMSSITVSEEDFILPVYKGELEKGYQFGWEISGEK 2016  
QY 61 DAGYVINSKDTFIRPVFKKIEKKEENKPTFDVSKKKNQOVNHSQLNESHKEDLOR 120  
DB 2017 DAGYVINSKDTFIRPVFKKIEKKEENKPTFDVSKKKNQOVNHSQLNESHKEDLOR 2076  
QY 121 BEHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 154  
DB 2077 EDHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 2110  
RESULT 5  
OSCPK8  
ID 08CPK8 PRELIMINARY; PRT; 775 AA.  
AC 08CPK8;  
DT 01-MAR-2003 (TRENBLREL. 23, Created)  
DT 01-MAR-2003 (TRENBLREL. 23, last sequence update)  
DT 01-MAR-2003 (TRENBLREL. 23, last annotation update)  
DE Penicillin-binding protein 1.  
GN S80856.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxId=1282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,

RA Chen Z., Wen Y.;  
RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AE016746; AAO04453.1; -.  
KW Complete proteome.  
SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;  
Query Match 14.8%; Score 118; DB 16; Length 775;  
Best Local Similarity 27.2%; Pred. No. 0.69;  
Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;  
QY 1 DTGEVSELKPHRVVTYVTONGKEMSSITVSEEDFILPVYKGELEKGYQFGW--- 51  
DB 615 DSVNAQSLKP-----ITIGNGKQIKQOVSQSGTKVLPKSKWMLMTDGLTMP-DMTGWTKE 669  
QY 52 EISGE-----GKKDGYVIN--LSKDTFIRPVFKKIEKKEENKPTFDVS---KK 98  
DB 670 DVLAFEDLTIKIKVSTKNGFVYNQISKQIK-----NKKIETVLSAEDT 716  
QY 99 KDNQVNHNSQLNESHKEDLOREHSQKSDSTKDVATATVLDKNNISSKSTNN 151  
DB 717 DDDDEKTEDSDSKSKDKXADEHSHNTSSSTKN-----DKSNADSGNDDSD 763  
RESULT 6  
OSCP15  
ID 09SP15 PRELIMINARY; PRT; 361 AA.  
AC 09SP15;  
DT 01-DEC-2001 (TRENBLREL. 19, Created)  
DT 01-DEC-2001 (TRENBLREL. 19, last sequence update)  
DT 01-OCT-2002 (TRENBLREL. 22, last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxId=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVO.  
RA Hisaeda H., Saul A., Long C.A., Miller L.H., Stowers A.W.;  
RT "Merozoite Surface Protein 3 and Protection Against Malaria in Aotus  
RT Monkeys."  
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AY044180; AAK94780.1; -.  
FT NON\_TER 1  
FT NON\_TER 361  
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;  
Query Match 14.1%; Score 112.5; DB 5; Length 361;  
Best Local Similarity 22.4%; Pred. No. 0.78;  
Matches 43; Conservative 33; Mismatches 55; Indels 61; Gaps 8;  
QY 1 DTGEVSELKPHRVVTYVTONGKEMSSITVSEEDFILPVYKGELEKGYQ----- 47  
DB 147 ETGE-----RNSRNNFYTTKTE-----YAGKVEKDYERAKANAYQKQNAV 187  
QY 48 -----FD-----GWEISGF--EGKKDAG-----YVINSKDTFIRPVFKKIEKKEEN 89  
DB 188 LKAKEASSDYIILGWEPGGVPEHKKENMLSHLVSSKDKENSKENDVDVLD--KEBA 246  
QY 90 KPTFDVSKKKNQOVNHSQLNESHKEDLOREHSQKSDSTKDVATATVLDKX----- 141  
DB 247 EETEEELKEKKEETSEHISEDEEEEEEKKEENDKKKEQKQKQSNENNDDKQME 306  
QY 142 --NISKSTNN 151  
DB 307 AQNLSKQNNN 318  
RESULT 7  
OSUC64  
ID 09UC64 PRELIMINARY; PRT; 379 AA.  
AC 09UC64;  
DT 01-MAY-2000 (TRENBLREL. 13, Created)  
DT 01-MAY-2000 (TRENBLREL. 13, last sequence update)

DT 01-OCT-2002 (Tremblrel. 22, last annotation update)  
DE Polymorphic antigen.  
GN MSP-3.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FCJ/HN;  
RA Li X.R., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;  
RT "Sequence of Plasmodium falciparum secreted polymorphic antigen  
RT gene."  
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AF188190; AAF04099.1; -  
SQ SEQUENCE 379 AA; 43316 MW; C152A54E1F9D5F25 CRC64;  
Query Match 14.1%; Score 112.5; DB 5; Length 379;  
Best Local Similarity 23.6%; Pred. No. 0.82;  
Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;  
QY 9 KPHRVTVTIQNGKMSSTIVSEDF-----ILPYKGLKGYQFD-GWEISGF- 56  
DB 171 KSRINLFSRKTKEVAYQV--EKDYERAKNAVQANQAVLKAKKASSYDYLIGMEFGGV 228  
QY 57 -EGKQDAG-----YVINLSKDTPIKPVFKIEKEEENKPTPDVSKKKDNPVNSQIN 110  
DB 229 PEHKKEENMLSHLYVSSKXKKNISKENDVDLDE-KEEBAEETEEELKEKNEETESIS 287  
QY 111 -----ESHKEDLQREHSQKSDSTKYDTATVLDKNISKSTNN 151  
DB 288 EDEEEEBEKEEENEKKEQKQKQENNDQKDMA-----QNLISKQNNN 336  
RESULT 8  
ID 025706 PRELIMINARY; PRT; 379 AA.  
AC 025706;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, last annotation update)  
DE Polymorphic antigen.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CS12;  
RA MEDLINE=98156743; PubMed=9497029;  
RA McColl D.J., Anders R.F.;  
RT "Conservation of structural motifs and antigenic diversity in the  
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
RL Mol. Biochem. Parasitol. 90:21-31(1997).  
DR EMBL; U08852; AAC47831.1; -  
SQ SEQUENCE 379 AA; 43302 MW; ABF9D54E1ED91A24 CRC64;  
Query Match 14.0%; Score 111.5; DB 5; Length 379;  
Best Local Similarity 23.6%; Pred. No. 0.98;  
Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;  
QY 9 KPHRVTVTIQNGKMSSTIVSEDF-----ILPYKGLKGYQFD-GWEISGF- 56  
DB 171 KSRINLFSRKTKEVAYQV--EKDYERAKNAVQANQAVLKAKKASSYDYLIGMEFGGV 228  
QY 57 -EGKQDAG-----YVINLSKDTPIKPVFKIEKEEENKPTPDVSKKKDNPVNSQIN 110  
DB 229 PEHKKEENMLSHLYVSSKXKKNISKENDVDLDE-KEEBAEETEEELKEKNEETESIS 287  
QY 111 -----ESHKEDLQREHSQKSDSTKYDTATVLDKNISKSTNN 151  
DB 288 EDEEEEBEKEEENEKKEQKQKQENNDQKDMA-----QNLISKQNNN 336  
RESULT 9

Q25705  
ID 025705 PRELIMINARY; PRT; 379 AA.  
AC 025705;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, last annotation update)  
DT 01-OCT-2002 (Tremblrel. 22, last annotation update)  
DE Polymorphic antigen.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K1;  
RA MEDLINE=98156743; PubMed=9497029;  
RA McColl D.J., Anders R.F.;  
RT "Conservation of structural motifs and antigenic diversity in the  
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
RL Mol. Biochem. Parasitol. 90:21-31(1997).  
DR EMBL; U08851; AAC47831.1; -  
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;  
Query Match 13.8%; Score 110.5; DB 5; Length 379;  
Best Local Similarity 22.1%; Pred. No. 1.2;  
Matches 43; Conservative 35; Mismatches 48; Indels 69; Gaps 9;  
QY 1 DTGEVSELKPHRVTVTIQNGKMSSTIVSEDFILPYKGLKGYQFD----- 47  
DB 167 ETGE-----RNSRNNFTYTKKE-----YAGKYKDYERAKNAVQANQAV 207  
QY 48 -----FD-----GWEISGF-EGKQDAG-----YVINLSKDTPIKPVFKIEKEEEN 89  
DB 208 LKAKEASSYDYLIGMEFGGVPEHKKEENMLSHLYVSSKXKKNISKENDVDLDE-KEEBA 266  
QY 90 KPTFPVSKKKDNPVNSQIN-----ESHKEDLQREHSQKSDSTKYDTAT 136  
DB 267 EETEEEBEKEEENEETESISEDEEBEKEEENEKKEQKQKQENNDQKDMA- 325  
QY 137 VLKKNISKSTNN 151  
DB 326 -----QNLISKQNNN 336  
RESULT 10  
ID 090G0 PRELIMINARY; PRT; 346 AA.  
AC 090G0;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
DT 01-OCT-2002 (Tremblrel. 22, last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
GN MSP3.  
OS Plasmodium reichenowii.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5854;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20416497; PubMed=10960178;  
RA Okemu D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in  
RT Plasmodium reichenowii and Plasmodium falciparum.";  
RL Mol. Biochem. Parasitol. 109:185-188(2000).  
DR EMBL; AJ252286; CAB65754.1; -  
FT NON\_TER 1 346  
FT NON\_TER 1 346  
SQ SEQUENCE 346 AA; 39127 MW; A804B96BDFAF010 CRC64;  
Query Match 13.8%; Score 110; DB 5; Length 346;  
Best Local Similarity 26.2%; Pred. No. 1.1;  
Matches 34; Conservative 26; Mismatches 54; Indels 16; Gaps 5;  
QY 37 VKKGLKGYQFD-GWEISGF-EGKQDAG-----YVINLSKDTPIKPVFKIEKEEEN 88  
DB 184 VLKAKEASSYDYLIGMEFGGVPEHKKEENMLSHLYVSSKXKKNISKENDVDLDE-KEEBA 242

QY 89 NKPTDVSKKKDPOVNHSHKEDLQREHSOKSDSTKDYATV-----LQKN 141  
 Db 243 ABEETGEQLEKNEKEETSEIINEDEQEBEKEEKEENKKEQAKQSDNQKEDMEQ 302  
 QY 142 NISSKSTNN 151  
 Db 303 NLISKQNNN 312

RESULT 11  
 Q26019 PRELIMINARY; PRT; 380 AA.  
 AC Q26019; 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Polymorphic antigen precursor.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FC27;  
 RX MEDLINE=95198774; PubMed=7891748;  
 RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,  
 RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.,  
 RA "Molecular variation in a novel polymorphic antigen associated with  
 RT Plasmodium falciparum merozoites".  
 RL Mol. Biochem. Parasitol. 68:53-67(1994).  
 DR EMBL; L07944; AAC09378.1; -.  
 KM Signal.  
 FT CHAIN 1 25 POTENTIAL.  
 FT SIGNAL 26 380 POLYMORPHIC ANTIGEN.  
 SQ SEQUENCE 380 AA; 43290 MW; 0986CA1393094CA2 CRC64;

Query Match 13.6%; Score 109; DB 5; Length 380;  
 Best Local Similarity 23.5%; Pred. No. 1.5;  
 Matches 42; Conservative 34; Mismatches 67; Indels 36; Gaps 8;

QY 2 TGEVSELKRVTVITIQNGKMSSTIVSEDFILPVY-----KGELEKGYQD-CWE 52  
 Db 166 TSETPE-KPSRINLPSRKTEYAEKAKNAYEKANAYOKANAYLKKAKASSYDYLQWE 224  
 QY 53 ISGF--EGKQAG-----VINLSKDTFKPVKKIEEKEENKPTFVSKKNPQVN 105  
 Db 225 FGCGVPEHKKEENMLSHLYSSKQKENIKENDVDLDE-KEEBAETSEEELKEKKEET 283

QY 106 HSQLN-----ESHRRKEDLQREHSOKSDSTDYATVLDKNNISSKSTNN 151  
 Db 284 ESEISDEEEEBEKEEENKKEQEKQESNNNDQKKQWEA-----QNLISKQNNN 337

RESULT 12  
 Q81436 PRELIMINARY; PRT; 3008 AA.  
 AC Q81436; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PFE0325W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=36329;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RP Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,  
 RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.,  
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22255708; PubMed=12368867;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harrie B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooke K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark E., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Fellwell T., Goble A., Goodhead I., Gilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphrey S., Jagels K., James K.D., Johnson D., Kerhornou A.,  
 RA Knights A., Kontorlov B., Kyes S., Larke N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Ruter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
 RL Nature 419:527-531(2002).  
 DR EMBL; AL929351; CAD51431.1; -.  
 KM Hypothetical protein.  
 SQ SEQUENCE 3008 AA; 356023 MW; 60BCBBE15C59984 CRC64;

Query Match 13.6%; Score 109; DB 5; Length 3008;  
 Best Local Similarity 32.4%; Pred. No. 14;  
 Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;

QY 64 YVINLSK---DTEFKVPFKIEEKEENKPTFVSKKNPQVNHQ---LNSHRKE 116  
 Db 2310 YDIELSKIEKFGASIGPFT-EEKKKEENK--EVNKKKEENKKEENKKEVNVKE 2366

QY 117 DLQREH-----SOKSDSTKDYATVLDKNNISSK-----STNNPNK 154  
 Db 2367 ENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENK 2414

RESULT 13  
 Q77355 PRELIMINARY; PRT; 600 AA.  
 AC Q77355; 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Hypothetical 71.7 kDa protein.  
 GN PFC0465C, MAL3P4.20.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=99376085; PubMed=10448855;  
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,  
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Fellwell T.,  
 RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,  
 RA Horrocks P., Jagels K., Jasal B., Kyes S., McLean J., Moule S.,  
 RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,  
 RA Ruter S., Skelton J., Squares R., Squares S., Sulston J.E.,  
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;  
 RT "The complete nucleotide sequence of chromosome 3 of Plasmodium  
 RT falciparum".  
 RL Nature 400:532-538(1999).  
 RL EMBL; AL008970; CA15610.2; -.  
 DR InterPro; IPR002483; PMI.  
 DR Pfam; PFO1480; PMI; 1.  
 DR SMART; SM00311; PMI; 1.  
 KM Hypothetical protein.

SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;  
 Query Match 13.6%; Score 108.5; DB 5; Length 600;  
 Best Local Similarity 29.3%; Pred. No. 2.7;  
 Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;  
 QY 27 IYSEEDFILPVY-----KGELEKGYQPDGWEISGFEKK-----DAGVINLSKDTFIKPY 77  
 DB 60 ILGFEDDILYETISQKSKSKK--DGEEDKYLAKKLKLTGFIKSKSDITIEL 116  
 QY 78 FKXI--EKKKE-----ENKPTFVSK-KKDNQVNSQLNE-----SHK 115  
 DB 117 LELLINBEKKKEHIADTLANK--TNDIKKYNENINENIVNENKDISKDKEHVSHQ 175  
 QY 116 E-----DLQEEH-----SQKSDTK---DVTATVLDKNNISSKSTN 150  
 DB 176 EHNINNVNLKEKEKYTDIQRDKRHRKSLQSKSDYKRPFPNRKTSIR-SLSNRRYDE 234  
 QY 151 NPK 154  
 DB 235 KTK 238  
 RESULT 14  
 Q90784 PRELIMINARY; PRT; 1038 AA.  
 AC Q90784;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Claustrin.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NC NCB1\_TaxID=9031;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RC Cole G.J.;  
 RN Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RP SEQUENCE OF 1-451 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=94157526; PubMed=7906711;  
 RT Burg M.A., Cole G.J.;  
 RT "Claustrin, an adhesive neural keratan sulfate proteoglycan, is  
 RT structurally related to MAP1B."  
 RT J. Neurobiol. 25:1-22(1994).  
 DR EMBL; X67778; CAA47988.1;--  
 SQ SEQUENCE 1038 AA; 117112 MW; 213D694A5B510927 CRC64;  
 Query Match 13.6%; Score 108.5; DB 13; Length 1038;  
 Best Local Similarity 28.6%; Pred. No. 4.9;  
 Matches 34; Conservative 25; Mismatches 43; Indels 17; Gaps 4;  
 QY 21 KEMSTIVSEEDFILPVYKLEKGYQPDGWEISGFEKKDAGVINLSKDTFIKPYK- 79  
 DB 595 KPEKTIIVAKDV-----TKKEQLGKSEKSEKQKQVKKPVTKSEKSEKKA 646  
 QY 80 KIEKKEEENKPFVDYKKKDNQVNSQLNEHREKDLQRE-----EHSQKSDTKKV 133  
 DB 647 KPEKDEKPKPEKSEKKEKPLI--KKEKPKEDIKKEVKEKKEKKEKKEV 702  
 RESULT 15  
 Q25995 PRELIMINARY; PRT; 354 AA.  
 AC Q25995;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Antigen.

OS Plasmodium falciparum (Isolate NF54).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NC NCB1\_TaxID=5843;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NF54;  
 RX MEDLINE=95198774; PubMed=7891748;  
 RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,  
 RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;  
 RT "Molecular variation in a novel polymorphic antigen associated with  
 RT Plasmodium falciparum merozoites."  
 RL Mol. Biochem. Parasitol. 68:53-67(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NF54;  
 RX MEDLINE=98156743; PubMed=9497029;  
 RA McColl D.J., Anders R.F.;  
 RT "Conservation of structural motifs and antigenic diversity in the  
 RT Plasmodium falciparum merozoite surface protein-3 (MSP-3)."  
 RL Mol. Biochem. Parasitol. 90:21-31(1997).  
 DR EMBL; L28825; AAC09377.1;--  
 SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;  
 Query Match 13.5%; Score 108; DB 5; Length 354;  
 Best Local Similarity 23.2%; Pred. No. 1.7;  
 Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;  
 QY 17 IONGKEMSTIVSEEDFILPVYKLEKGYQPDGWEISGF--EGKKDAG-----YINIS 69  
 DB 181 VIKAKKAS-----YDYL-----GWEFGGVGEHKKKEEMLSHLYVSSKD 221  
 QY 70 KDTFIKPVFKIEKKKE-----ENKPTFVSKKKDNQVNSQLNESHRKE 116  
 DB 222 KENISKENDVDLDEKEEBEETBEELKEKEEETSESEDEEBEKEEKEEENDKK 281  
 QY 117 DLQEEHQSQSDTKDVTATVLDKNNISSKSTNN 151  
 DB 282 EOEKQSNENNQKKDME-----QNLISKNNNN 311  
 RESULT 16  
 Q81U55 PRELIMINARY; PRT; 354 AA.  
 AC Q81U55;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Merozoite surface protein 3.  
 GN PPI0.0345.  
 OS Plasmodium falciparum (Isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NC NCB1\_TaxID=36329;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berrihan M., Hyman R.W.,  
 RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angiolini S.,  
 RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Fraser C.M., Barrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum."  
 RL Nature 419:498-511(2002).  
 DR EMBL; AE014834; AAN35542.1;--  
 SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;  
 Query Match 13.5%; Score 108; DB 5; Length 354;  
 Best Local Similarity 23.2%; Pred. No. 1.7;

Matches 36; Conservative 29; Mismatches 46; Indels 44; Gaps 6;

QY 17 IONGKMSSTIVSEDFILPYKGELEKGYQPDGWEISGF--EGKKDAG----YVNLIS 69  
 DB 181 VIKAKEASS-----VDYIL-----GMEFGGVEHEKKENMLSHLYVSSKD 221  
 QY 70 KOTFIKVPFKIEEKKE-----ENKPTFVSKKKKDNQVNHSLNSHRE 116  
 DB 222 KENISKENDVDLDEKEEAEETBEERLEKNEEETSESTSEDBEEBEEBEEBEEBNDKK 281  
 QY 117 DLQREHSQKSDSTKDVATVLDKNNISSKSTNN 151  
 DB 282 EOEKQSNENNOKKDMA-----QNLISKNNNN 311

RESULT 17  
 Q9FJK9 PRELIMINARY; PRT; 470 AA..  
 AC Q9FJK9;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE GblAPF20218.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosida II; Brassicales; Brassicaceae; Arabidopsie.  
 OX NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=99087489; PubMed=9872454;  
 RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
 RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
 RT physically assigned P1 and TAC clones.";  
 RL DNA Ref. 5:297-308 (1998).  
 RL EMBL; AB015468; BAB10694.1; -  
 SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72B35AC54 CRC64;

Query Match 13.4%; Score 107.5; DB 10; Length 470;  
 Best Local Similarity 20.1%; Pred. No. 2.5;  
 Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

QY 9 KPRVTVTIQNGKMSSTIVSEDFILPYKGELEKGYQPDGWEISGFE-----GRK 60  
 DB 82 RNRVTVDITQNNNSNGSK-----YVODLARIYDE-EATGSGSAGRIDHPNK 129  
 QY 61 DAGVYINLSKOTFIKVPFKIEEKKEENKPTFVSKKKDN----- 101  
 DB 130 NGGITETAKAPENSPIEETSHRVDDNKRINNQKFTAKSSENAVSRVPGADHRAEVMGK 189  
 QY 102 PQVNHSLNE-----SHREKDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPK 154  
 DB 190 PMENDQVQTESAEKSHRKEVNTYSEKPRDQGVAKTEAKDKRKEKEKETSINK 248

RESULT 18  
 Q815F3 PRELIMINARY; PRT; 829 AA..  
 AC Q815F3;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PFL1275C.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_Taxid=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;

RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Anguilo S.,  
 RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Frazer C.M., Barrrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum.";  
 RL Nature 419:498-511 (2002).  
 DR EMBL; AB014848; AAN36341.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 829 AA; 98816 MW; EF2675E301B2CE93 CRC64;

Query Match 13.4%; Score 107; DB 5; Length 829;  
 Best Local Similarity 24.8%; Pred. No. 5;  
 Matches 36; Conservative 30; Mismatches 43; Indels 36; Gaps 6;

QY 26 TIVSEDFILPYKGELEKGYQPDGWEISGFEKKDAGVYINLSKOTFIKVPFKIEBK 85  
 DB 519 TLMTRDVLVDVFPFPMYKKTHLN-----KKETLFNFSLN-----FREIEKVK 561  
 QY 86 EEENKPT-----FDVSKKKDN-PQVNHSLNSH-RKEDLQREHSQKSDS 129  
 DB 562 RDKKKGTINKNDNAEEYMLKIKKKKKNEENNNTLDSNIRKENNKLVHEH--DNS 618  
 QY 130 TKDVATVATVLDKNNISSKSTNNPK 154  
 DB 619 LKQEOIITINDKQVIEHTKIYDNQKK 643

RESULT 19  
 Q96229 PRELIMINARY; PRT; 951 AA..  
 AC Q96229;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PFB0680W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_Taxid=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=99021743; PubMed=9804551;  
 RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,  
 RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,  
 RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Petrea M.,  
 RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
 RA Frazer C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
 RT "Chromosome 2 sequence of the human malaria parasite Plasmodium  
 RT falciparum.";  
 RL Science 282:1126-1132 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W., K.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Anguilo S.,  
 RA Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Frazer C.M., Barrrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum.";





RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laevo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";  
 RA Science 287:2185-2195 (2000).  
 RL [12]  
 RN SEQUENCE FROM N.A.  
 RP Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Barton J., An H., Baldwin D., Banton J., Beeson K.Y., Buesam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsetti V., Doup L.E., Doyle C., Drenek D., Farfan D.,  
 RA Ferreira S., Fries E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson C.A., Nunco J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phonananong S., Peltman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RA "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochownik S.E., Smith C.D.,  
 RA Tupy J.L., Begeman C., Bereman B., Carlson J.W., Celinker S.E.,  
 RA Clamp N., Drysdale R., Emmert D., Fries E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Seale S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.V., Lewis S.E.;  
 RA "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RA Flybase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [16]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-20564328; PubMed-10956655;  
 RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
 RA O'Neill K.T., Pocht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
 RA Friedman P.A.;  
 RA "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved  
 RT Isoform of Asph Missing the Catalytic Domain Share Exons with  
 RT Junction.";  
 RL J. Biol. Chem. 275:39543-39554 (2000).  
 DR EMBL, AE003808; AAF58063.1; -;  
 DR EMBL, AF289494; AAC40807.1; -;  
 DR Flybase; Fgn0034075; Asph.  
 SQ SEQUENCE 382 AA; 43287 MW; 605EC03AEBFC6E8B CRC64;

Query Match 12.7% Score 101.5; DB 5; Length 382;  
 Best Local Similarity 24.5%; Pred. No. 5.5;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 QY 31 EDFILPVYKGLKNGYQPDG-----EISGPEKQAGYI-----NLKPTPIK 75  
 DB 78 EDDLTPLESEFSK-VFDGAVDEHDEHDHDDHDPDEHDHDDHDEDEDE 135  
 QY 76 PVFKIIEKKEENKPT-----FDVSKKKNPOVNSQLNESHKPELQREHSQKSDS 129  
 DB 136 PLTELELELELEEEPEPEDEPADDEEYEDDEENNA--GENITAEDEEEEDND 193  
 QY 130 TKDVTATVLDNNISKST 148  
 DB 194 EGVFAIVEATTEATTEAT 212  
 RESULT 23  
 ID Q9V719 PRELIMINARY; PRT; 556 AA.  
 AC Q9V719;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE CG8421 protein.  
 GN ASPH OR CG8421 OR CG18658.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephyrididae; Drosophilidae; Drosophila.  
 CX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Barton R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Abmayant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,  
 RA Borovka D., Botchan D.A., Bouck J., Brokstein P., Brotlier P.,  
 RA Burris K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laevo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RA "The genome sequence of *Drosophila melanogaster*.";

RL Science 287:2185-2195(2000).

[2]

RP SEQUENCE FROM N.A.

RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferreria S., Frise E., Galle R.F., Gary N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,

RA Phouenavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,

RT "Sequencing of Drosophila melanogaster genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[3]

RP SEQUENCE FROM N.A.

RA Mistra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceiniker S.E.,

RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,

RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield B.,

RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall J.C., Lewis S.E.,

RT Annotation of Drosophila melanogaster genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[4]

RP SEQUENCE FROM N.A.

RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[5]

RP SEQUENCE FROM N.A.

RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AEO03808: AAF58064.2; -

DR FlyBase: FBgn0034075; Asph.

SQ SEQUENCE 556 AA; 63144 MW; B420980CBDC357A CRC64;

Query Match 12.7%; Score 101.5; DB 5; Length 556;

Best Local Similarity 24.5%; Pred. No. 8.3;

Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 31 EDLILPVYKGLKGYQPDGM-----EISGFEKKDAGYVI-----NISKPTFK 75

DB 78 EDIDTLSESRFSK--VFQGVNDEHDEHGHVQSPSGEALDDHDEHDDHDEDEE 135

QY 76 PVKPKLEKKKEKKPKP-----FVSKKKDNQVNHVSQLESKREDDLOREHSQKSDS 129

DB 136 PLTEELBEELBEELPEDEDEPADEBEDEDEENNA--GENITADAEBEEDNDND 193

QY 130 TKDVTATVLDKNNISKST 148

DB 194 EGTVEATVEATTEATTEAT 212

RESULT 24

Q9G082 PRELIMINARY; PRT; 785 AA.

AC Q9G082; 01-MAR-2001 (TEMBLrel. 16, Created)

DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)

DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA).

GN ASPH OR CG8421 OR CG18658

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidae; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

LN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=20564328; PubMed=10956665;

RX Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,

RA O'Neil K.T., Pocht R.J., Scully M.S., Hollis J.M., Hollis G.F.,

RA Friedman P.A.

RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved

RT Isoform of Asph Missing the Catalytic Domain Share Exons with

RT Junction."

RL J. Biol. Chem. 275:39543-39554(2000).

[2]

RP SEQUENCE FROM N.A.

RA STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne C., Miklos G.L.G.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Baldwin D.,

RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankuch C., Beasley E.M.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhanderi D., Bolshakov S.,

RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Mikulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Nusslein D.R., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacleb J.M.,

RA Palazolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter J., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).

[3]

RP SEQUENCE FROM N.A.

RA Ceiniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferreria S., Frise E., Galle R.F., Gary N.S., George R.A.,

RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,

RA Phouenavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,

RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,

RT "Sequencing of Drosophila melanogaster genome."

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

[4]

RP SEQUENCE FROM N.A.

RA Mistra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,

RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Ceiniker S.E.,

RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Milburn G., Richter J., Russo S.,  
 RA Seale S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
 RT "Annotation of Drosophila melanogaster genome."  
 RT Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Gelinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.,  
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Flybase;  
 RU Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF289493; AAC40806.1; -;  
 DR EMBL; AB003808; AAM70947.1; -;  
 DR Flybase; FBgn0034075; Asph.  
 DR InterPro; IPR001440; TPR.  
 DR InterPro; IPR006025; Zn\_PROTEASE.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 DR SEQUENCE 785 AA; 89843 MW; 30A8DFCD6836F7F1 CRC64;

Query Match 12.7%; Score 101.5; DB 5; Length 785;  
 Best Local Similarity 24.5%; Pred. No. 12;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;

QY 31 EEFILPVYKGELEKGYQFPGW-----ELSGREGKKDAGYV-----NLKDTFLK 75  
 DB 78 ELDLPPLSRPSRK--VFQGWVDEHRDEHDGVDVPESEALDDHDDHDDHDEDEDE 135  
 QY 76 PVFKLEEKKEENKPT-----FDVSKKKNQVHNSOLNSHREKDLQREHSOKSDS 129  
 DB 136 PTEELIEELIEEETEDERADDEYEDDEENNA--GENTLADAEKEEEDNDND 193  
 QY 130 TKDVTATVLDKNNISKST 148  
 DB 194 EGTVEATVEATTEATTEAT 212

RESULT 25

Q9NFV9 PRELIMINARY; PRT; 329 AA.  
 AC Q9NFV9;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Merzoitte surface protein 3 (Fragment).  
 GN MSP3.  
 OS Plasmodium falciparum (isolate 7G8).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=57266;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=7G8;  
 RX MEDLINE=20416497; PubMed=10960178;  
 RA Okenu D.M.N., Thomas A.W., Conway D.J.;  
 RT "Allelic lineages of the merzoitte surface protein 3 gene in  
 RT Plasmodium reichenowi and Plasmodium falciparum."  
 RT Mol. Biochem. Parasitol. 109:185-188 (2000).  
 DR EMBL; AJ252287; CAB85901.1; -;  
 DR NON\_TER 1  
 FT NON\_TER 1  
 FT NON\_TER 329  
 SQ SEQUENCE 329 AA; 36916 MW; CSB045DB5E21A159 CRC64;

Query Match 12.6%; Score 101; DB 5; Length 329;  
 Best Local Similarity 25.3%; Pred. No. 5.1;  
 Matches 38; Conservative 25; Mismatches 47; Indels 40; Gaps 7;

QY 17 IONGKEMSSITVSEEDFILPVYKGELEKGYQFPGWISGFG--EGKKDAG-----YINIS 69  
 DB 171 VUKAKKAS-----VDYIL-----GMFPGGVGVPEHKKKEENMLSHLYVSKD 211  
 QY 70 KOTFLKPVKTIKKEEENKPTFDVSKKKNQVHNSOLNSHREKDLQREHS----- 123

DB 212 KENISKENDVLDK-KEBEAEETEBEELKEEKEETESISEDEBEEKEEKEQAKE 270  
 QY 124 --SKSDSTKDVATVLDKNNISKSTNN 151  
 DB 271 QSNENNDDKXIMEA-----QVLISKNNNN 295

RESULT 26

Q9PPL5 PRELIMINARY; PRT; 312 AA.  
 ID Q9PPL5  
 AC Q9PPL5;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Putative membrane protein.  
 GN CU0692C.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;  
 OC Campylobacteraceae; Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIC 11168;  
 RX MEDLINE=20150912; PubMed=10688204;  
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,  
 RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,  
 RA Jags K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,  
 RA Whitehead S., Barrall B.G.;  
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
 RT reveals hypervariable sequences."  
 RL Nature 403:665-668 (2000).  
 DR EMBL; AL139076; CAB72966.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 312 AA; 37221 MW; 0004FA7836A741E8 CRC64;

Query Match 12.5%; Score 100; DB 16; Length 312;  
 Best Local Similarity 25.3%; Pred. No. 5.7;  
 Matches 41; Conservative 32; Mismatches 61; Indels 28; Gaps 8;

QY 17 IONGKEMSSITVSEEDFILPVYK-----GELEKGYQFPGWISGFGKKDAGYV 65  
 DB 24 INQKPLDDDLRDEISSDILARRKKTTPKPLEEDLEYESKTKTKSNLYKED---L 80  
 QY 66 INL---SDDTFLPVYKTIKKEEENKPT--FDVSKKKNP-----QVHNSOLNSHREK 115  
 DB 81 INVKLEEKQSLAKKIFSKKERRKKNKTNFLFSKKANEIKNIQTQIOITKSNQA 140  
 QY 116 EDLQREHSOKSDSTKDV--TATVLDKNNISK--STNNNN 153  
 DB 141 TTQTKQEKELTNLSIEKIQTETKIQPLILEKLDVKNQPN 182

RESULT 27

O14712 PRELIMINARY; PRT; 375 AA.  
 ID O14712  
 AC O14712;  
 DT 01-JAN-1998 (TREMblrel. 05, Created)  
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Cell cycle progression restoration 8 protein.  
 GN CPR8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98043401; PubMed=9383053;  
 RA Edwards M.C., Liegeois N., Horecka J., Depinho R.A., Sprague G.F. Jr.,  
 RA Tyers M., Elledge S.J.;  
 RT "Human Cpr (cell cycle progression restoration) genes impart a far-



Best Local Similarity 24.7%; Pred. No. 48;  
Matches 36; Conservative 24; Mismatches 60; Indels 26; Gaps 5;

QY 26 TIVSEDFLPVYKGELEKGYQDGEISGF-----EKKQAGVYINLSKDT- 72  
DB 216 TDYSEDEIYAKYIDKSDNSGYDKSKLINTSNINMLNTYKKNQVNH--SMSSNTI 273  
QY 73 -----FIVPVFKIIEKKKEENKPTFDVSKKQDNQVNHSQLNESHKREDLQREHSQKS 127  
DB 274 QODLSFIHSSINKYKKEKKEKKN--YDKKKSSNTNDSYNTITONDPKKNQKSEFVDNN 332  
QY 128 DSTKQVATVLDKNNISSKSTNNPN 153  
DB 333 NGRND-----HNKNNELEQVYNNPN 353

RESULT 31  
0813A0  
ID 0813A0 PRELIMINARY; PRT; 2563 AA.  
AC 0813A0;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DE E1-R2 ATPase/hydrolase, putative.  
CN PF10240C.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22255708; PubMed=12368867;  
RA Hall N., Pain A., Bettiman M., Churcher C., Harris B., Harris D.,  
Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
Felwell T., Goble A., Goodhead I., Gilliam R., Hamlin N., Hance Z.,  
Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
Humphrey S., Jagels K., Kyes K.D., Johnson D., Kethornou A.,  
Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
Seeger K., Sharp S., Smith R., Squares R., Stevens K.,  
Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
Sulston J.E., Craig A., Newbold C., Barrell B.G;  
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
RL Nature 419:527-531(2002).  
DR EMBL; AL929355; CAD51734.1; --  
KW Hydrolase.  
SQ SEQUENCE 2563 AA; 298805 MW; 3F9613243D26F8F1 CRC64;

Query Match 12.4%; Score 99; DB 5; Length 2563;  
Best Local Similarity 29.4%; Pred. No. 68;  
Matches 35; Conservative 18; Mismatches 60; Indels 6; Gaps 4;

QY 40 GELKGYQFDG-WEISGFEKQDAGVYINLSKDTPIKPVFKIIEKKKEENKPTFDVSKK 98  
DB 856 GEEKNG-NIDGIYILKQKNKKDKMIGEBENKDNFSEKKEKSDNENSEIDKNNVYLKR 914  
QY 99 KD--NPOVNSQLNESHKREDLQREH-SQKSDSTKYDTATVLDKNNISSKSTNNPN 153  
DB 915 KEKHNSNDEDNINDSVLKENVKEKEIHSSSSNEOSDYLKKEKKDKNNISVDNNEN 973

RESULT 32  
081226  
ID 081226 PRELIMINARY; PRT; 238 AA.  
AC 081226;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Hypothetical protein.

GN PF10765M.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22255708; PubMed=12368867;  
RA Hall N., Pain A., Bettiman M., Churcher C., Harris B., Harris D.,  
Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
Felwell T., Goble A., Goodhead I., Gilliam R., Hamlin N., Hance Z.,  
Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
Humphrey S., Jagels K., Kyes K.D., Johnson D., Kethornou A.,  
Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
Seeger K., Sharp S., Smith R., Squares R., Stevens K.,  
Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
Sulston J.E., Craig A., Newbold C., Barrell B.G;  
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
RL Nature 419:527-531(2002).  
DR EMBL; AL929356; CAD51839.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 238 AA; 28481 MW; DA175A2B5A109B49 CRC64;

Query Match 12.3%; Score 98.5; DB 5; Length 238;  
Best Local Similarity 28.1%; Pred. No. 55;  
Matches 48; Conservative 25; Mismatches 63; Indels 35; Gaps 10;

QY 4 EYSELKPRVYVYTONGKEMSTIVSEDFLPVYKGELEKGYQDGEISGF--FEKK 60  
DB 42 KTOETKED-----ENKSNNNVVSSENF-MKIYKNLIKQNYHYEKKQKQENIESEK 95  
QY 61 DA-----GVYINLSKDTPIKPVFKIIEKKKEENKPTFDVSKKQDN--POV--NH--S 107  
DB 96 KATSKKKIKKLNFKLKKNNLLEKTVK--KAFELAKTVPSPLKKYTVPIPSNNHIS 152  
QY 108 QLNESHKRED-----LQREHSQKSDSTKYDTATVLDKNNISSKSTNNPN 153  
DB 153 QNDHKKKQDKKIKKIKLLEKSNQDNN-----NINNKNNQKKDPTNNIN 198

RESULT 33  
09OV00  
ID 09OV00 PRELIMINARY; PRT; 157 AA.  
AC 09OV00;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE CG31958 protein.  
GN CG31958 OR CG10022.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydriidae; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutcliffe G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.H.C., Blazer R.G., Chame M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Domes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merklow G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,  
 RA Palazolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Bonzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.B., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouenavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Mistra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Bertan B., Carlson J.W., Celniker S.E.,  
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Seearle S.M.J., Smith E., Shu S., Smutnak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.D., Lewis S.E.;  
 RT "Annotation of *Drosophila melanogaster* genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Playbase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR HMBL, AAC03578; AAF51062.2; -;  
 DR HMBL, P02593; ICTR.  
 DR FLYbase; FBgn0051958; CG31958.  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF00036; ehand; 3.  
 DR ProDom; PD000012; EF-hand; 1.  
 DR SMART; SM00054; EFh; 3.  
 DR PROSITE; PS00018; EF HAND; 1.  
 KW Calcium; Calcium-binding.

SQ SEQUENCE 157 AA; 18263 MW; 543F0480E119DEC1 CRC64;  
 Query Match 12.3%; Score 98; DB 5; Length 157;  
 Best Local Similarity 22.7%; Pred. No. 3.8;  
 Matches 34; Conservative 29; Mismatches 53; Indels 34; Gaps 5;  
 QY 23 MSSTVSEEDFLIPYKGLKGYQFDGMEISGFEKKDAGYI-----NLSKDTF 73  
 DB 1 MELSVSEDDILKNITY-SLDR-----DNGALTSKELGWIVIALGRPNESIAKEEF 52  
 QY 74 IPEVFKIIE-KKEENKPTPVSKKDN-----PQVHSLNESHKRE 116  
 DB 53 CNVIIRKMDTKREBELRDAFVFPKKNNGYISTELRAVFNALGEKDELLEIMIRY 112  
 QY 117 DIQREHSOKSDSTQDVATVADKNNISK 146  
 DB 113 DLDQDHNIFESNNMTPIINVTAVK 142  
 RESULT 34  
 ID 0813J9 PRELIMINARY; PRT; 4524 AA.  
 AC 0813J9;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 GN PFE1325W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_Taxid:36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Devlin K., Baker S., Davies P., Mungall K., Berriman M., Pain A.,  
 RA Hall N., Bowman S., Churcher C., Quail M., Barrett B.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:2255708; PubMed:12368867;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,  
 RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphrey S., Jagels K., James K.D., Johnson D., Kerhovan A.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Leonard N.,  
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinovitch E.,  
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Sultson J.E., Craig A., Newbold C., Barrett B.G;  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
 RL Nature 419:527-531(2002).  
 DR EMBL; AL929354; CAD51629.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 4524 AA; 525287 MW; 542206173C29567A CRC64;  
 Query Match 12.3%; Score 98; DB 5; Length 4524;  
 Best Local Similarity 24.8%; Pred. No. 1.5e+02;  
 Matches 40; Conservative 26; Mismatches 75; Indels 20; Gaps 6;  
 QY 2 TGEVSELKPHRYVTI-----ONGKMSSTVSEEDFLIPYKGLKGYQFDGMEI 53  
 DB 3832 TGEFDDIKPEVSESLISNNNNNNKKSKNNIVKNSMVVHYINQY---YFNSIRY 3888  
 QY 54 SGFEKGKAGYINISKDTFIKPVKKEEENKPTPVSKKDNPOVHSLNESH 113  
 DB 3889 SNFNLSIDLN-MINLNKN-SITITDKIKSKCHNRMSISIKKIRSVYIKKS 3946  
 QY 114 RKEDLQREHSOKSDSTQDVATVADKNNISKSTTNPNK 154



DB 3947 MNIIQKET-EKKQDTJKEV-----KNKLKKSKDANKKNK 3980

## RESULT 35

081HW3 PRELIMINARY; PRT; 449 AA.

ID Q81HW3  
AC Q81HW3  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN PF11\_0413.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
NCBI\_TaxId=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RC MEDLINE=22255705; PubMed=12368864;

RA Gardner M.J., Hall N., Fung E., White O., Bertman M., Hyman R.W.,  
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angiuoli S.,  
RA Perera M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA Venner J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrel B.;  
RT "Genome sequence of the human malaria parasite Plasmodium falciparum."  
RL Nature 419:498-511 (2002).  
DR EMBL, AF014842; AAN3596.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 449 AA; 52812 MW; 85B62272D6257C68 CRC64;

Query Match 12.2%; Score 97.5; DB 5; Length 449;  
Best Local Similarity 25.1%; Pred. No. 13;  
Matches 42; Conservative 28; Mismatches 48; Indels 49; Gaps 7;

QY 21 KEMSTTIVSEDFILPVY-----KGELEKGYOPFGWIEGFEKKDAGVYINLSKDTF 73  
DB 55 KELTITILKEK--ILILVNEQNEKKKKKKKQFQSD---ISKKKSKYADLQKSSSKNN 109  
QY 74 IKPVEFKIE-----EKKEE-----NKPTFDV-----SKKQNPQVN 105  
DB 110 LKKKKKIIESEBETINSDDEEEBEGYQKKQKKQKNSVSTLSLEKKKKKKRSESEN 169  
QY 106 HSQINESHKEDLQREHSGQSDSTKQVATVLDKNNISSKSTYNNP 152  
DB 170 NNDVNEEYDEDDDEQEEEEEES-----LSKKSKKKKSTSTP 207

## RESULT 36

095S93 PRELIMINARY; PRT; 556 AA.

ID 095S93  
AC 095S93;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE GM05229P.  
GN ASPH OR CG8421 OR CG18658.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydriidae; Drosophilidae; Drosophila.  
NCBI\_TaxId=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,

RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;  
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL, AY060905; AAL28453.1; --  
DR Flybase; FBgn0034075; Asph.  
SQ SEQUENCE 556 AA; 63089 MW; 95D82EAC57D11FE8 CRC64;

Query Match 12.2%; Score 97.5; DB 5; Length 556;  
Best Local Similarity 24.5%; Pred. No. 17;  
Matches 34; Conservative 28; Mismatches 52; Indels 25; Gaps 5;

QY 31 EDFILPVYKGELEKGYOPDGW-----EISGFEKKDAGVYI-----NLSKDTITK 75  
DB 78 EDLPTLSESRFSK--VPDGWVDEHDEHDGVDVQEPGEALDDHDEHDHDEDEBE 135  
QY 76 PVFKIEEKKEKKEKKEKPT-----PVSKKKQNPQVNSQNLNESHKEDLQREHSGQSDS 129  
DB 136 PLTELELELELEBEETPEDEPPADEEYEBDEBENNA--GENITABDAVEEBEEDND 193  
QY 130 TKQVATVLDKNNISSKST 148  
DB 194 EGTVAIVATTEATTEAT 212

## RESULT 37

09K5S1 PRELIMINARY; PRT; 614 AA.

ID 09K5S1  
AC 09K5S1;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hypothetical protein BH4017.  
GN BH4017.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
NCBI\_TaxId=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RC MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasaawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis."  
RL Nucleic Acids Res. 28:4317-4331 (2000).  
DR EMBL, AP001520; BAB07736.1; --  
DR InterPro; IPR006829; Transposase 30.  
DR Pfam; PF04740; Transposase 30; 1--  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 614 AA; 70704 MW; 2E3FDE6C74E0294C CRC64;

Query Match 12.2%; Score 97.5; DB 16; Length 614;  
Best Local Similarity 27.3%; Pred. No. 18;  
Matches 41; Conservative 26; Mismatches 44; Indels 39; Gaps 8;

QY 6 SELKPRVYTVIIONKEMSTTIVSE-----EDFILPVYKGELEKGYOPDGW-----EISG 55  
DB 386 SEPKPE--TYLTQTAIQW-TPIVNEYSPTREBEF-----ARAHQLDGWADVSKVY 435  
QY 56 FEGKKDAGVYINLSKDTITKPVFKIEKK--EENKPTFDVSKK-----DNPQVNSQL 109  
DB 436 FAGNNIASQL-----GKIEEKQODKNTNNTTDFPKKEVNVVOSPIKSTAL 482  
QY 110 NESHKEDLQREHSGQSDSTKQVATVLD 139  
DB 483 GKVIIGIVDLGRKHKHGEDLRRLSKSQIE 512

## RESULT 38

093424 PRELIMINARY; PRT; 385 AA.

ID 093424  
AC 093424; P90801;  
DT 01-JAN-1999 (TREMBLrel. 09, Created)



DT 01-JAN-1999 (T-EMBLrel. 09, last sequence update)  
DT 01-MAR-2003 (T-EMBLrel. 23, last annotation update)  
DE Hypothetical glycine-rich 37.0 kDa protein E02A10.2 in chromosome V precursor.  
GN E02A10.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Pelodierinae; Caenorhabditis.  
OC NCB1\_Taxid=6239;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Thomas K.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP REVISIONS.  
RC STRAIN=BRISTOL N2;  
RA Jones S.J.M.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; 281053; CAB02877.1; -.  
DR HSSP; P10968; 2CWC; CE09116.  
DR WormPep; E02A10.2; EGShell.  
DR InterPro; IPR002952; EGShell.  
DR PRINTS; PR01228; EGShell.  
KM Hypothetical protein; Signal; Transmembrane.  
FT SIGNAL 1 17  
FT CHAIN 18 385  
FT TRANSMEM 165 185  
FT DOMAIN 23 190  
SQ SEQUENCE 385 AA; 36963 MW; 32AB25AE9B5073FE CRC64;  
  
Query Match 12.1%; Score 97; DB 5; Length 385;  
Best Local Similarity 29.6%; Pred. No. 12;  
Matches 42; Conservative 24; Mismatches 48; Indels 28; Gaps 7;  
  
QY 1 DNGEVELKPHRYTVIQ--NGKMSSTIVSEDFILPYKGLKGYQFDGWEISGFSG 58  
DB 256 ETDDFPSLQENVHCTILINIKVARK--NEEDKKEEKKKEEKEEVEKKE--EDB 310  
QY 59 KIDAGVINLSKDTFIPVKFKLIEKKKEENKPTFVSKKKONPQVNSQLNESHRKEDL 118  
DB 311 KQDE-----EP--KKEEKKKEEKKKE--EVEKKE-----EKKDEPKKEE 349  
QY 119 QREHSQKSDSTKDTATVLDK 140  
DB 350 KKEEKKKEEKEEKEEKEE 371  
  
RESULT 39  
Q81ST1 PRELIMINARY; PRT; 558 AA.  
AC Q81ST1;  
DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
DT 01-MAR-2003 (T-EMBLrel. 23, last sequence update)  
DT 01-MAR-2003 (T-EMBLrel. 23, last annotation update)  
DE Hypothetical protein.  
GN PFL060W.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OC NCB1\_Taxid=36329;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RA MEDLINE=22255705; PubMed=1236864;  
RA Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W., Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyse S., Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S., Pereira M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B., Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A., McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C., Venter J.C., Carnucci D.J., Hoffman S.L., Newbold C., Davis R.W.,

RA Fraser C.M., Barrell B.;  
RT "Genome sequence of the human malaria parasite Plasmodium falciparum";  
RL Nature 419:498-511 (2002).  
DR EMBL; AE014846; AAN36209.1; -.  
KM Hypothetical protein.  
SQ SEQUENCE 558 AA; 66487 MW; 2AD029EA69F90D3 CRC64;  
  
Query Match 12.1%; Score 97; DB 5; Length 558;  
Best Local Similarity 22.4%; Pred. No. 18;  
Matches 35; Conservative 31; Mismatches 64; Indels 26; Gaps 4;  
  
QY 18 QNGKMSSTIVSEDFILPYKGLKGYQFDGWEISGFSGKDAGYVINLSKDTFIRV 77  
DB 41 QNGKRYIYNLSKKKNGGIIKGGSKHGHF-----INNYKKRNVKYLKTKYKPSISF 96  
QY 78 FKIEKKKEENKPTFVSKKKONPQVNSQLNE-----SHRKE 118  
DB 97 FKK-KDKNDKNDK--DKTHMKDNGNLNKKKEGEGIKDKEKEKEKILIKDKKDKDE 153  
QY 119 QREHSQKSDSTKDTATVLDKNNISSKSTTNPK 154  
DB 154 KKKKELQKEKKTKLTVLKKNNKNNNNNNNNK 189  
  
RESULT 40  
Q9STN4 PRELIMINARY; PRT; 988 AA.  
AC Q9STN4;  
DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
DT 01-MAY-2000 (T-EMBLrel. 13, last sequence update)  
DT 01-OCT-2002 (T-EMBLrel. 22, last annotation update)  
DE Hypothetical 109.0 kDa protein.  
GN T28D5.30 OR AT4G08340.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; OC eustids II; Brassicales; Brassicaceae; Arabidopsis.  
OC NCB1\_Taxid=3702;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Bevan M., Lennard N., Quail M., Harris B., Rajandream M.A., RA Barrell B.G., Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K., RA Scheller C.;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
RN (3)  
RP SEQUENCE FROM N.A.  
RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G., RA Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN (4)  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL109819; CAB52556.1; -.  
DR EMBL; AL161511; CAB77959.1; -.  
DR InterPro; IPR001760; Opain.  
DR InterPro; IPR003653; SUMO protease.  
DR Pfam; PF02302; Peptidase\_C48; 1.  
DR PROSITE; PS00238; OPSIN; 1.  
DR PROSITE; PS06000; UTP\_PROTEASE; 1.  
KM Hypothetical protein.  
SQ SEQUENCE 988 AA; 108955 MW; 01C518587D460EAD CRC64;  
  
Query Match 12.1%; Score 97; DB 10; Length 988;  
Best Local Similarity 22.1%; Pred. No. 34;  
Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;  
  
QY 19 NKGKMSSTIVSEDFILPYKGLKGY-----QFDGWEISGFSGKDAGYVINLSKDTF 73

Db 383 NGRQNSNVQSSVDIELSYTTPDKVPSGVGLNVSERDIVEVEVEDVRSAGGLSPNVQORDN- 441  
Qy 74 IKPVFKIEEKE-----EENKPTFDVSKKQNPQVNHSQLNESHREKDLQREE 122  
Db 442 VEPVGDDVRSAGDMSPNPSAANNVREGPATFDIMESEDNPGRDVNAFPMEDHIRSEVQLSP 501  
Qy 123 HSQKSDSTKDYT--ATVLDKNNISSKSTNNPNK 154  
Db 502 HVL---GAKDVTVDSDPTDKVGVNDVTDASDPT 532

Search completed: February 10, 2004, 10:57:07  
Job time : 28.2661 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 37.3096 Seconds  
(without alignments)  
697.707 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773  
Perfect score: 848  
Sequence: 1 TTVEFLINKDTGEVSELEK.....ATVLDKNNISKSTTNPNK 164

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues  
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	848	100.0	773	22 AAB48343	S. pneumoniae Sp13
2	848	100.0	2120	21 AAB48343	S. pneumoniae Sp13
3	848	100.0	2140	24 AAB48343	S. pneumoniae type
4	848	100.0	2140	24 AAB48343	S. pneumoniae type
5	615	72.5	117	19 AAB55096	Streptococcus pneu
6	615	72.5	117	19 AAB55096	Streptococcus pneu
7	119	14.0	746	23 AAB39023	S. epidermidis ope
8	107.5	12.7	484	21 AAB47777	Staphylococcus epi
9	106	12.5	665	21 AAB18278	Arabisopsis thalia
					Plasmodium falcipa

10	101.5	12.0	564	22 ABB61977	Drosophila melanog
11	99.5	11.7	1384	24 ABP55413	Human MDP2-22 prot
12	99.5	11.7	1404	24 ABP55393	Human MDP2-2 prote
13	97.5	11.5	2519	22 ABB16636	Novel human diagn
14	96.5	11.4	408	21 AAG37134	Arabidopsis thalia
15	96.5	11.4	456	21 AAG37133	Arabidopsis thalia
16	96.5	11.4	476	21 AAG37132	Arabidopsis thalia
17	96.5	11.4	2060	23 AAE20967	Staphylococcus lug
18	96	11.3	2500	21 AAB18272	Plasmodium falcipa
19	93.5	11.0	645	24 ABB19106	Pathogen specific
20	93.5	11.0	654	24 ABB56879	Staphylococcus epi
21	92.5	10.9	645	24 ABB18979	Pathogen specific
22	92	10.8	299	22 AAG82397	S. epidermidis ope
23	92	10.8	309	23 ABB40822	Staphylococcus epi
24	91	10.7	442	23 ABB38188	Staphylococcus epi
25	91	10.7	902	23 ABB40312	Staphylococcus epi
26	90.5	10.7	209	21 AAB42897	Human ORF261
27	90.5	10.7	635	23 ABB73209	Candida albicans e
28	90.5	10.7	757	21 AAY44364	Human cell cycle r
29	90.5	10.7	758	21 AAB53319	Human colon cancer
30	90.5	10.7	801	23 ABB68963	Human polypeptide
31	90.5	10.7	922	22 AAB58023	Shrimp white spot
32	90	10.6	1183	22 ABB58769	Drosophila melanog
33	89.5	10.6	402	20 AAY03190	S. aureus trigger
34	89.5	10.6	511	20 AAY35091	Chlamydia pneumoni
35	89.5	10.6	516	23 ABB73570	Candida albicans e
36	89.5	10.6	525	20 AAB03189	S. aureus trigger
37	89.5	10.6	1408	22 ABB58704	Drosophila melanog
38	89	10.5	296	23 ABB93341	Herbicidally activ
39	88.5	10.4	281	22 ABB64828	Drosophila melanog
40	88.5	10.4	1141	22 AAG85008	Shrimp white spot
41	88	10.4	225	23 ABB73992	Candida albicans e
42	88	10.4	258	22 AAB94584	Human protein sequ
43	88	10.4	817	22 AAM79318	Human protein SEQ
44	88	10.4	817	22 AAM79318	Human protein SEQ
45	88	10.4	2515	22 ABB12281	Human secreted pro

## ALIGNMENTS

RESULT 1  
ID AAB48343 standard; Protein; 773 AA.  
AAB48343;  
20-APR-2001 (first entry)  
S. pneumoniae Sp130 polypeptide.  
Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;  
bronchial; lung; blood; infection; immune response; immunotherapy;  
antibacterial; auditory; vaccine.  
OS Streptococcus pneumoniae.  
PN WO200076540-A2.  
PD 21-DEC-2000.  
PF 09-JUN-2000; 2000MO-US15925.  
PR 10-JUN-1999; 99US-0138453.  
PA (MED1-) MED IMMUNE INC.  
PI Adamou JE, Choi GH;  
WPI; 2001-112197/12.  
DR N-PSDB; AAC84742.  
XX New vaccines comprising Sp128 or Sp130 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by  
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
PT blood infections -

PS Claim 8; Page 51-54; 54pp; English.

XX The invention relates to novel immunogenic polypeptides, Sp128 and Sp130  
XX from *S. pneumoniae*. Vaccines comprising the polypeptides are useful for  
CC the treatment and prevention of pneumococcal infections, particularly  
CC infections caused by *Streptococcus*, such as otitis media, nasopharyngeal,  
CC bronchial, lung or blood infections. The antigens are used as immunogenic  
CC agents to stimulate an immune response. The antisera and antibodies may  
CC also be used in diagnosing and treating pneumococcal infections.  
CC Recombinant polypeptides serve as a mechanism for stimulating production  
CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
CC as reagents in other processes such as affinity chromatography. The  
CC present sequence represents the *S. pneumoniae* Sp130 polypeptide.

SQ Sequence 773 AA;

Query Match 100.0%; Score 848; DB 22; Length 773;

Best Local Similarity 100.0%; Pred. No. 1.1e-74;

Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEPILNKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDPILPVYKGELEKGYQFDG 60

DB 610 TTVKEPILNKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDPILPVYKGELEKGYQFDG 669

QY 61 WEISGEGKKDAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSKKKDNQVNHSQLN 120

DB 670 WEISGEGKKDAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSKKKDNQVNHSQLN 729

QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164

DB 730 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 773

RESULT 2

AA81710

ID AAY81710 standard; Protein; 2120 AA.

AC AAY81710;

DT 02-JUN-2000 (first entry)

DE Streptococcus pneumoniae protein sequence ID3.

XX Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;

KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;

KW kidney disease; diabetes; immunosuppressive disorder; otitis media;

XX pneumococcal septicemia; sinusitis; meningitis; therapy.

OS Streptococcus pneumoniae.

XX WO200006738-A2.

PN 10-FEB-2000.

PD 27-JUL-1999; 99WO-GB02452.

PF 27-JUL-1998; 98GB-0016336.

PR 19-MAR-1999; 99US-0125329.

XX (MICR-) MICROBIAL TECHNIQS LTD.

PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;

XX WPI: 2000-195301/17.

DR N-PSDB; AA291806.

XX Streptococcal proteins and polynucleotides useful for diagnosis,

PT treatment and prophylaxis of bacterial infections -

PS Claim 2; Page 41-42; 76pp; English.

XX This sequence represents a *Streptococcus pneumoniae* protein of the  
CC invention. The proteins (or their homologues, derivatives and/or  
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
CC compositions comprising the proteins are useful as vaccines and also in  
CC diagnostic assays. The sequences are useful for the detection or  
CC diagnosis of *S. pneumoniae* infection, by contacting a sample to be tested  
CC with them. Agents capable of antagonising, inhibiting or interfering with  
CC the function or expression of the protein or polypeptide are useful in  
CC medical compositions in the treatment or prophylaxis of *S. pneumoniae*  
CC infection. As the sequences can be used to treat *S. pneumoniae* infection,  
CC they can be used to treat bacterial pneumonia, which has high rates in  
CC young children, the elderly, and in patients with predisposing conditions  
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
CC or with immunosuppressive disorders, especially AIDS. They can also be  
CC used to treat pneumococcal septicemia, otitis media, sinusitis, and  
CC meningitis.

SQ Sequence 2120 AA;

Query Match 100.0%; Score 848; DB 21; Length 2120;

Best Local Similarity 100.0%; Pred. No. 4.3e-74;

Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKEPILNKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDPILPVYKGELEKGYQFDG 60

DB 1923 TTVKEPILNKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDPILPVYKGELEKGYQFDG 1982

QY 61 WEISGEGKKDAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSKKKDNQVNHSQLN 120

DB 1983 WEISGEGKKDAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSKKKDNQVNHSQLN 2042

QY 121 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164

DB 2043 ESHRKEDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2086

RESULT 3

AB01020

ID AB01020 standard; Protein; 2140 AA.

AC AB01020;

DT 11-FEB-2003 (first entry)

DE *S. pneumoniae* type 4 strain protein from coding region #590.

XX Bacterial meningitis; pneumonia; sepsis; otitis media;

KW ear infection; antiinflammatory; antibacterial; immunostimulant;

KW auditory; respiratory; gene therapy; vaccine.

XX Streptococcus pneumoniae type 4 strain.

OS WO200277021-A2.

PN 03-OCT-2002.

PD 27-MAR-2002; 2002WO-IB02163.

PF 27-MAR-2001; 2001GB-0007658.

XX (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Masignani V, Tettelin H, Frazer C;

XX WPI: 2003-040579/03.

DR N-PSDB; ABX06302.

XX New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,

PT useful as medicaments for treating or preventing a disease or infection

due to streptococcus bacteria, such as pneumonia, sepsis, otitis media

PT or ear infection -

PS Claim 1; SEQ ID No 1180; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment,  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
CC AB556454. Also included are an antibody which binds one of the  
CC proteins, treating a patient by administering the protein, DNA or  
CC antibody (in a composition), a kit comprising first and second primers,  
CC which are the nucleic acid cited above or fragments between nucleotides  
CC 8-100 of a sequence not defined in the specification, for amplifying a  
CC target sequence contained within a Streptococcus nucleic acid sequence,  
CC where the first primer is substantially complementary to the target  
CC sequence and the second primer is substantially complementary to the  
CC complement of the target sequence, and where the parts of the primers  
CC having substantial complementarity define the termini of the target  
CC sequence to be amplified, assay comprising contacting a test compound  
CC with the protein, and determining whether the test compound binds to the  
CC protein and a Streptococcus pneumoniae bacterium, where one or more  
CC genes encoding the proteins has been rendered inactive. The proteins,  
CC nucleic acid molecules, antibody and compositions are useful as  
CC medicaments for treating or preventing a disease or infection due to  
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,  
CC sepsis, otitis media or ear infection. They are also useful in developing  
CC vaccines, diagnostics and antibiotics. The methods are useful for  
CC identifying immunodominant proteins. The present sequence is one of  
CC the 2469 proteins expressed by the identified coding regions from the  
CC genomic sequence.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 2140 AA;

Query Match 100.0%; Score 848; DB 24; Length 2140;  
Best Local Similarity 100.0%; Pred. No. 4.4e-74;  
Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVEFIANKDTGSEVSELPKRVVTYTIQNGKMSSTIVSEDFILPYVKGLEKGYODG 60

DB 1943 TTVEFIANKDTGSEVSELPKRVVTYTIQNGKMSSTIVSEDFILPYVKGLEKGYODG 2002

QY 61 WEISGFEKGDAGYVINSKDTFIKVPFKIEEKEENKPTFDVSKKDNPOVNSQLN 120

DB 2003 WEISGFEKGDAGYVINSKDTFIKVPFKIEEKEENKPTFDVSKKDNPOVNSQLN 2062

QY 121 ESHRKEDLQREHSOKSDSTKDVTATVLDKNNISSKSTNNPNK 164

DB 2063 ESHRKEDLQREHSOKSDSTKDVTATVLDKNNISSKSTNNPNK 2106

RESULT 4

ID AAM55096 standard; Protein; 117 AA.

AC AAM55096;

DT 02-OCT-1998 (first entry)

DE Streptococcus pneumoniae SP0043 protein.

KM Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;

KW detection; pneumonia; otitis media; meningitis.

OS Streptococcus pneumoniae.

PN MO9818930-A2.

PD 07-MAY-1998.

XX 30-OCT-1997; 97MO-US19422.

PR 31-OCT-1996; 96US-0029960.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

DR WPI, 1998-272224/24.

DR N-PSDB; AAV27357.

PT Nucleic acid encoding antigenic peptide(s) from Streptococcus

PT pneumoniae - or their epitope-containing fragments, useful in

PT protective or therapeutic vaccines, and for diagnosis

PS Claim 11; Page 62; 118pp; English.

CC The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.

SQ Sequence 117 AA;

Query Match 72.5%; Score 615; DB 19; Length 117;  
Best Local Similarity 100.0%; Pred. No. 7.3e-53;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 YKGELEKGYQDFGWEISGFEKGDAGYVINSKDTFIKVPFKIEEKEENKPTFDVSK 107

DB 1 YKGELEKGYQDFGWEISGFEKGDAGYVINSKDTFIKVPFKIEEKEENKPTFDVSK 60

QY 108 KKNPQVNSQLNESHKREDLQREHSOKSDSTKDVTATVLDKNNISSKSTNNPNK 164

DB 61 KKNPQVNSQLNESHKREDLQREHSOKSDSTKDVTATVLDKNNISSKSTNNPNK 117

RESULT 5

ID ABP54590 standard; Protein; 117 AA.

AC ABP54590;

DT 04-SEP-2002 (first entry)

DE S. pneumoniae SP043 protein sequence SEQ ID NO:68.

KM Streptococcus pneumoniae; epitope; vaccine; antigenic protein;

KW antibacterial; Streptococcal infection; detection.

OS Streptococcus pneumoniae.

PN US2002061545-A1.

PD 23-MAY-2002.

PR 22-JAN-2001; 2001US-0765272.

PR 30-OCT-1997; 97US-0961083.

PA (CHOI/) CHOI G H.

PA (KUNS/) KUNSCH C A.

```
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR,
PI Rosen CA;
XX WPI; 2002-479261/51.
DR N-PSDB; AB084825.
XX
PT New Streptococcus pneumoniae antigens, useful for detecting
PT Streptococcus and for preventing or attenuating disease caused by
PT Streptococcus infection -
XX
XX Claim 11; Page 29; 70pp; English.
XX
XX AB084792 to AB084904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
CC The S. pneumoniae antigens have antibacterial activity and can be
CC used in vaccines. The S. pneumoniae antigens can also be used to
CC prevent or attenuate a Streptococcal infection in an animal. The
CC polynucleotides encoding the S. pneumoniae antigens can be used to
CC detect Streptococcus nucleic acids. AB084905 to AB085130 represent
CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
CC which are used in an example from the present invention.
XX
SQ Sequence 117 AA;
XX
Query Match 72.5%; Score 615; DB 23; Length 117;
Best Local Similarity 100.0%; Pred. No. 7.3e-53;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 48 YKGELEKGYQPDGMEISGFEKKKAGYVINTLSKDTFIKPVFKKIEKKEENKPTDVS 107
DB 1 YKGELEKGYQPDGMEISGFEKKKAGYVINTLSKDTFIKPVFKKIEKKEENKPTDVS 60
XX
QY 108 KKDNPQVNHSQLNESHRKEDLOREHESOKSDSTKYDTATVLDKNNISSKSTNN 164
DB 61 KKDNPQVNHSQLNESHRKEDLOREHESOKSDSTKYDTATVLDKNNISSKSTNNPK 117
XX
RESULT 6
AA081779
ID AA081779 standard; Protein; 746 AA.
XX
AC AA081779;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX KW vaccination; endocarditis.
XX
XX Staphylococcus epidermidis.
XX OS
XX WO200134809-A2.
XX PN
XX 17-MAY-2001.
XX PD
XX 09-NOV-2000; 2000WO-US30782.
XX PF
XX 09-NOV-1999; 99US-0164258.
XX PR
XX (GLAX ) GLAXO GROUP LTD.
XX PA
XX Kimerly MJ;
XX PI
XX WPI; 2001-316495/33.
XX DR N-PSDB; AAH52629.
XX
```

```
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 18; Page 208; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (i) encoding polypeptides
CC (ii), given in AA081454 to AA083120, from Staphylococcus epidermidis.
CC (i) and (ii) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (i) may be used to produce the
CC S. epidermidis polypeptides (ii) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (ii) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 746 AA;
XX
Query Match 14.0%; Score 119; DB 22; Length 746;
Best Local Similarity 27.0%; Pred. No. 0.0081;
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;
XX
QY 10 KDTGEVSEIKPRVVTITONGKEMSTYSEEDFILPYVK-----GELEKGYQPDG 61
DB 585 EDSVNAQSLKP-----ITIGNGKQIKQGSYSGTKVLPHSKWLMTDGETLP-DMTGWTK 639
XX
QY 62 -EISGFE-----GKQAGYVINTLSKDTFIKPVFKKIEKKEENKPTDVS 107
DB 640 EYVLAPELITKLKVTGKNGFTNOSISKQIYK-----NKDKIEVLSAED 686
XX
QY 108 KKDNPQVNHSQLNESHRKEDLOREHESOKSDSTKYDTATVLDKNNISSKSTNN 161
DB 687 TDDQEKTEDESDNKSXKDKADEHDSNTSSSTKN-----DKSNADSKNDSD 734
XX
RESULT 7
ABP39023
ID ABP39023 standard; Protein; 778 AA.
XX
AC ABP39023;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX KW antibacterial; gene therapy.
XX
XX Staphylococcus epidermidis.
XX OS
XX US6380370-B1.
XX PN
XX 30-APR-2002.
XX PD
XX 13-AUG-1998; 98US-0134001.
XX PF
XX 14-AUG-1997; 97US-055779P.
XX PR
XX 08-NOV-1997; 97US-064964P.
XX PA
XX (GENO-) GENOME THERAPEUTICS CORP.
XX PI
XX Doucette-Stamm LA, Bush D;
XX
```



DR WPI: 2002-381255/41.  
DR N-PSDB; ABN91568.  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
PS Disclosure; SEQ ID 3868; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP3124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 778 AA;  
Query Match 14.0%; Score 119; DB 23; Length 778;  
Best Local Similarity 27.0%; Pred. No. 0.0086;  
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
Qy 10 KDTGEVSELKPRVTVITQNGKMSSTVSEEDFLPYVK-----GELEKGYQPDGW-- 61  
Db 617 EDSVNAQSLKP-----ITIGNGKQIKQSVKSGTKVLPKSHKVMMLMTDGLTWP-DWTGWTK 671  
Qy 62 -EISGFE-----GKDAQGYIN--LSKDTFIKVFKEKKEENKPTFDVS---K 107  
Db 672 EDVLAFFEDLTIKIVSTKNGFTVNTQSIKQGIK-----NDKTEVLSLAE 718  
Qy 108 KKDQNVNHSQLESRKEDLQREHSQKSDTKDTATVLDKNNISKSTTN 161  
Db 719 TDDDEKTEDESDSKSKKDKADEHSHNTSSSTKA-----DKSNADSNDD 766  
RESULT 8  
AAG47777  
ID AAG47777 standard; Protein; 484 AA.  
XX  
AC AAG47777;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60255.  
XX  
KM protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
EN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142927.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.

```

PR    14-OCT-1999;      99US-0159637.      99US-0159638.
PR    14-OCT-1999;      99US-0159638.      99US-0159638.
PR    18-OCT-1999;      99US-0159584.      99US-0159584.
PR    21-OCT-1999;      99US-0160741.      99US-0160741.
PR    21-OCT-1999;      99US-0160767.      99US-0160767.
PR    21-OCT-1999;      99US-0160768.      99US-0160770.
PR    21-OCT-1999;      99US-0160770.      99US-0160814.
PR    21-OCT-1999;      99US-0160814.      99US-0160815.
PR    21-OCT-1999;      99US-0160980.      99US-0160980.
PR    22-OCT-1999;      99US-0160981.      99US-0160981.
PR    22-OCT-1999;      99US-0160989.      99US-0161404.
PR    25-OCT-1999;      99US-0161404.      99US-0161405.
PR    25-OCT-1999;      99US-0161405.      99US-0161406.
PR    25-OCT-1999;      99US-0161359.      99US-0161359.
PR    26-OCT-1999;      99US-0161360.      99US-0161360.
PR    26-OCT-1999;      99US-0161361.      99US-0161920.
PR    28-OCT-1999;      99US-0161920.      99US-0161993.
PR    28-OCT-1999;      99US-0161993.      99US-0161993.
PR    28-OCT-1999;      99US-0162142.      99US-0162142.
PR    29-OCT-1999;      99US-0162142.      99US-0162142.

Query Match          12.7%; Score 107.5; DB 21; Length 484;
Best Local Similarity 20.1%; Pred. No. 0.06; Mismatches 66; Indels 45; Gaps 5;
Matches   36; Conservative   32;

Oy     19 KPHRVTTVTONGKEMSSSTIVSEEDFILPVYKGLELGYQPDGWISGEF-----GKK 70
           :|::||::||::||::||::||::||::||::||::||::||::||::||::||:
Db     96 RENVTVDTVOQNNSNGESK-----YYODIARRIRYDE-FATGSOSQRIDHPNOK 143
Oy       71 DAGVINLSKDPTFKLPFPVKIEKEEENKPDPVSCKKN-----            111
           :::::||||:::::||::|::|::|::|::|::|::|::|::|::|::|::|
Db     144 NVGITKAFFESPLETESHRDVDNKRLNNQNFPAKSSENAAVRSGADHKRAEVWGK 203
           ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy     112 POUHSQLNE-----SRRKEDLOREHSQSQSDTKDVTALVDKNNISSKTNTNPDK 164
           |=====|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     204 PMEHRDVQRCHSAEKSRKRKENVTYSKRPDOBGVTKTEAKDKDRNKEKETESLNK 262
           =====|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 9
AAB18278 ID AAB18278 standard; Protein; 665 AA.
XX AC AAB18278;
XX DT 07-NOV-2000 (first entry)
XX DE Plasmodium falciparum chromosome 2 related protein SBQ ID NO:135.
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine; antimalarial; malaria; protozoacide; infection; insecticide.
XX OS Plasmodium falciparum.
XX PN WO200025728-A2.
XX PD 11-MAY-2000.
XX PE 05-NOV-1999; 99WO-US26796.
XX PR 05-NOV-1998; 98US-0107131.
XX PA (HOFF/) HOFFMAN S.
XX PA (CARU/) CARUCCI D.
XX PA (GAR/) GARDNER M.
XX PA (VENT/) VENTER J C.
XX PI Hoffman S., Carucci D., Gardner M., Venter JC;
XX WP: 2000-365347/31.
XX PT Proteins encoded by chromosome 2 of the human malarial parasite.
```

PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
PT diagnosis of P.falciparum infection -  
XX Disclousre; Page 321-322; 577pp; English.  
XX  
XX The present invention describes proteins and their fragments (I) encoded  
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
CC vaccines against P. falciparum infection comprising (I) or (II).  
CC (I) and (II) are useful for the development of vaccines against  
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
CC antibody raised to immunogens comprising the sequences of (I), are  
CC useful in the detection of infection with P. falciparum. Furthermore,  
CC (I) (especially when they are rifins or secreted or membrane proteins)  
CC can aid the identification of drugs to treat or prevent P. falciparum  
CC infection, or they can be used to identify drug resistance in  
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic lifecycle, and provide new targets for  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AAAY0078 to AAAY0287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.  
XX  
SQ Sequence 665 AA;  
Query Match 12.5%; Score 106; DB 21; Length 665;  
Best Local Similarity 24.3%; Pred. No. 0.13; Indels 56; Gaps 10;  
Matches 42; Conservative 35; Mismatches 40;  
OY 10 KDTGVESEIKPRVAV-VTIQNGKEMSVTVSEDEPILPVYKGLERKGYQFDGMEI--SGF 66  
DB 127 EEKNKINSDLRHQELNMQSK-----NEQDI-----NKNKGKQ---DLSNSVA 169  
OY 67 EGKQDAGVYINLSKDTFKVPFKIEKKE-----EENKPPD---VSKKQNP 112  
DB 170 EENKXP-----VEGVKELEKKEKEKISDHHKVENKKSDDHKVENKKSDDH 217  
OY 113 QVNHSQLNESHKEDLQR-EESHQKSDSTKDTATVLDKNNSKSTTNPNK 164  
DB 218 KVENKKSDDHKIEVKVEEHEDEBE-----DKKESSENKKNK 261  
RESULT 10  
ABB61977  
ID ABB61977 standard; Protein; 564 AA.  
XX  
XX ABB61977;  
AC  
XX  
XX 26-MAR-2002 (first entry)  
DT  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 12723.  
DE  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
XX Drosophila melanogaster.  
OS  
XX  
XX WO200171042-A2.  
PN  
XX  
XX 27-SEP-2001.  
PD  
XX  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX  
XX 11-JUL-2000; 2000US-0614150.  
PA  
XX  
XX (PEKE ) PE CORP NY.  
PI Venter JC, Adams M, Li PWD, Myers EW;

XX  
DR WPI; 2001-656860/75.  
DR N-PSDB; ABL06080.  
XX  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclousre; SEQ ID NO 12723; 21pp + Sequence listing; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins  
CC (AB57737-AB872072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 564 AA;  
Query Match 12.0%; Score 101.5; DB 22; Length 564;  
Best Local Similarity 24.5%; Pred. No. 0.29; Indels 25; Gaps 5;  
Matches 34; Conservative 29; Mismatches 51;  
OY 41 EDFILPVYKGLERKGYQFDGW-----EISGFEKQKQGYI-----NLKQTFIK 85  
DB 78 EDLDTPLSESRFSK--VFDGVDNDRHDDHNDVQPERGELADDDHDDHDDHDEDEE 135  
OY 86 PVFKIEEKEEENKPT-----FDVSKKQNPQVNHSQLNESHKEDLQREHSHQKSDS 139  
DB 136 PLTELELELEEBEPEDEPADDEYEDEDEENNA--GENTIADEEBEEREBEEDMD 193  
OY 140 TKDVTATVLDKNNSKST 158  
DB 194 EGTEATVTEATTEAT 212  
RESULT 11  
ABP55413  
ID ABP55413 standard; Protein; 1384 AA.  
XX  
XX ABP55413;  
AC  
XX  
XX 04-FEB-2003 (first entry)  
DT  
XX  
XX Human MDDT-22 protein SEQ ID NO:22.  
DE  
XX  
XX Human; MDDT; molecules for disease detection and treatment; anti-HIV;  
KW antiallergic; antiinflammatory; antianemic; antiparkinsonian; nocotropic;  
KW anticonvulsant; antiinfectivity; antiarteriosclerotic; antiasthmatic;  
KW immunosuppressive; antithyroid; cytostatic; hepatotropic; dermatological;  
KW antidiabetic; nephrotoxic; antitumor; thyromimetic; neuroprotective;  
KW osteopathic; antiarthritic; antiparasitic; antihelminthic; antipneumatic;  
KW uropathic; ophthalmological; antirheumatic; haemostatic; antibacterial;  
KW virucide; protozoacide; fungicide; gene therapy; cell proliferative;  
KW cancer; developmental disorder; neurological disorder; infection;  
KW reproductive disorder; autoimmune disorder; inflammatory disorder.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200278420-A2.  
PN  
XX  
XX 10-OCT-2002.  
PD  
XX  
XX 29-MAR-2002; 2002WO-US09809.  
PF  
XX  
XX 30-MAR-2001; 2001US-280387P.  
PR  
XX  
XX 05-APR-2001; 2001US-282335P.  
PA  
XX  
XX 13-APR-2001; 2001US-283663P.

PR 19-APR-2001; 2001US-285484P.  
PR 18-JAN-2002; 2002US-350702P.  
PR 25-JAN-2002; 2002US-351749P.  
XX  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
PI Lu DAM, Arvizu CS, Gandhi AR, Hafalia AJA, Ding L, Lu Y;  
PI Rankumar J, Swarnakar A, Tang YT, Yue H, Tran B, Lee SY;  
PI Warren BA, Nguyen DB, Thangavelu K, Yao MG, Elliott VS, Baughn MR;  
PI Emerling BM, Lal PG, Gietzen KJ, Becha SD, Marquis JP, Kable AE;  
XX MPI; 2003-058385/05.  
DR N-PSDB; AB083880.  
XX  
PT New human molecules for disease detection and treatment, useful for  
PT diagnosing, treating or preventing autoimmune or inflammatory disorders  
PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,  
PT cancer or hepatitis -  
XX  
PS Claim 1; Page 207-211; 238pp; English.  
XX  
XX The present invention describes 23 human molecules for disease detection  
CC and treatment (MDPT-1 to 23) (see ABP55392 to ABP55414). The human  
CC MDPT-1 to 23 proteins (I) are encoded by the sequences given in AB083859  
CC to AB083881. (I) can have various activities depending on the cells and  
CC tissues in which they are expressed. These activities include: anti-HIV;  
CC antiallergic; antiinflammatory; antianaemic; antiparkinsonian; nootropic;  
CC anticonvulsant; antifertility; antiatherosclerotic; antiasclerotic;  
CC immunosuppressive; antihypertrophic; cytoskeletal; hepatotropic; dermatological;  
CC antidiabetic; nephrotropic; antigout; thymomatous; neuroprotective;  
CC osteopathic; ophthalmological; antiparasitic; antihelminthic; antibacterial;  
CC uroepathic; ophthalmological; antiparasitic; haemostatic; antiparasitic;  
CC virucide; protozoacide; and fungicide. (I) and the polynucleotides  
CC encoding them can be used in gene therapy. (I), polynucleotides, agonists  
CC and antagonists from the present invention can be used for diagnosing,  
CC treating or preventing disorders associated with aberrant expression of  
CC MDPT, particularly cell proliferative (e.g. cancer), developmental  
CC disorders, neurological disorders, reproductive disorders, or  
CC autoimmune/inflammatory disorders, or viral, bacterial, fungal,  
CC parasitic, protozoal or helminthic infections. They are also useful in  
CC the assessment of the effects of exogenous compounds on the expression  
CC of nucleic acid and amino acid sequences of proteins associated with  
CC MDPT.  
XX  
XX  
SQ Sequence 1384 AA;  
XX  
XX Query Match 11.7%; Score 99.5; DB 24; Length 1384;  
XX Best Local Similarity 21.7%; Pred. No. 1.6;  
XX Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;  
XX  
QY 4 KEPIINKDQGEVSELEKPHVTVYITQNGKMSSTIVSEEDPILPVYKGELEKGYQF----- 58  
DB 1013 KEPIMLQNEQELISQLK-KEIERIQQMKMESYMKQDEQYINATQYKEADLDQCELRITRE 1071  
QY 59 -----DGEWISGEGKKDA-----GYVINSK----- 80  
DB 1072 QVONSTBELAARHQOVQORERIRLSSELEDMKQSKKDAAGNHLAELGASKREAH 1131  
QY 81 -----DTFKPVFKKLEEKKE-----ENKPTFDVSKKKNDQVNSQLSNESHRE 126  
DB 1132 LEARMQAEIKLSAEVESLKEAYHMEMISHQENHAKWTKIS--ADSGKSSVQQLNEQLERA 1189  
QY 127 DLQREHHSQKSDSTKVATVATVDKNNI 153  
DB 1190 KLELEB---ADPTVSNLHOQOVODRNEV 1213  
XX  
XX RESULT 12  
XX ID ABP55393 standard; protein; 1404 AA.  
XX AC ABP55393;  
XX

DT 04-FEB-2003 (first entry)  
XX  
DE Human MDPT-2 protein SEQ ID NO:2.  
XX  
XX Human; MDPT; molecules for disease detection and treatment; anti-HIV;  
KW antiallergic; antiinflammatory; antianaemic; antiparkinsonian; nootropic;  
KW anticonvulsant; antifertility; antiatherosclerotic; antiasclerotic;  
KW immunosuppressive; antihypertrophic; cytoskeletal; hepatotropic; dermatological;  
KW antidiabetic; nephrotropic; antigout; thymomatous; neuroprotective;  
KW osteopathic; antiparasitic; antihelminthic; antibacterial;  
KW uroepathic; ophthalmological; antiparasitic; haemostatic; antiparasitic;  
KW virucide; protozoacide; and fungicide; gene therapy; cell proliferative;  
KW cancer; developmental disorder; neurological disorder; infection;  
KW reproductive disorder; autoimmune disorder; inflammatory disorder.  
XX  
XX Homo sapiens.  
XX  
XX WO200278420-A2.  
XX  
XX 10-OCT-2002.  
XX  
XX 29-MAR-2002; 2002WO-US09809.  
XX  
XX 30-MAR-2001; 2001US-280387P.  
XX 05-APR-2001; 2001US-282335P.  
XX 13-APR-2001; 2001US-283663P.  
XX 19-APR-2001; 2001US-285484P.  
XX 18-JAN-2002; 2002US-350702P.  
XX 25-JAN-2002; 2002US-351749P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
PI Lu DAM, Arvizu CS, Gandhi AR, Hafalia AJA, Ding L, Lu Y;  
PI Rankumar J, Swarnakar A, Tang YT, Yue H, Tran B, Lee SY;  
PI Warren BA, Nguyen DB, Thangavelu K, Yao MG, Elliott VS, Baughn MR;  
PI Emerling BM, Lal PG, Gietzen KJ, Becha SD, Marquis JP, Kable AE;  
XX MPI; 2003-058385/05.  
DR N-PSDB; AB083880.  
XX  
DR New human molecules for disease detection and treatment, useful for  
PT diagnosing, treating or preventing autoimmune or inflammatory disorders  
PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,  
PT cancer or hepatitis -  
XX  
XX  
PS Claim 1; Page 172-175; 238pp; English.  
XX  
XX The present invention describes 23 human molecules for disease detection  
CC and treatment (MDPT-1 to 23) (see ABP55392 to ABP55414). The human  
CC MDPT-1 to 23 proteins (I) are encoded by the sequences given in AB083859  
CC to AB083881. (I) can have various activities depending on the cells and  
CC tissues in which they are expressed. These activities include: anti-HIV;  
CC antiallergic; antiinflammatory; antianaemic; antiparkinsonian; nootropic;  
CC anticonvulsant; antifertility; antiatherosclerotic; antiasclerotic;  
CC immunosuppressive; antihypertrophic; cytoskeletal; hepatotropic; dermatological;  
CC antidiabetic; nephrotropic; antigout; thymomatous; neuroprotective;  
CC osteopathic; antiparasitic; antihelminthic; antibacterial;  
CC uroepathic; ophthalmological; antiparasitic; haemostatic; antibacterial;  
CC virucide; protozoacide; and fungicide. (I) and the polynucleotides  
CC encoding them can be used in gene therapy. (I), polynucleotides, agonists  
CC and antagonists from the present invention can be used for diagnosing,  
CC treating or preventing disorders associated with aberrant expression of  
CC MDPT, particularly cell proliferative (e.g. cancer), developmental  
CC disorders, neurological disorders, reproductive disorders, or  
CC autoimmune/inflammatory disorders, or viral, bacterial, fungal,  
CC parasitic, protozoal or helminthic infections. They are also useful in  
CC the assessment of the effects of exogenous compounds on the expression  
CC of nucleic acid and amino acid sequences of proteins associated with  
CC MDPT.  
XX  
XX  
SQ Sequence 1404 AA;  
XX  
XX Query Match 11.7%; Score 99.5; DB 24; Length 1404;  
XX

Best Local Similarity 21.7%; Pred. No. 1.6;  
Matches 45; Conservative 32; Mismatches 67; Indels 63; Gaps 7;  
QY 4 KEFLINKDTGSEVSELPKPHVTVTQNGKMSSTIVSEEDFILPYKGELEKGYOPDGEI 58  
DB 1033 KEFLINKDTGSEVSELPKPHVTVTQNGKMSSTIVSEEDFILPYKGELEKGYOPDGEI 1091  
QY 59 -----DGEISGPEGKDA-----GVINLSK----- 80  
DB 1092 QVONSHTELEAKRHOQVQOREIERLSSELEDMKOLSKEDHGNHLELQASKVREAH 1151  
QY 81 -----DTFIKPVFKKIEKKK-----EENKPTFDVSKKKDNPQVNSQLNESHRE 126  
DB 1152 LEARMQAEIKKLSAEVESLKEAYVHMEMISHQENHAKWIS--ADSQKSVQQLNEQLKKA 1209  
QY 127 DQREHSGKSDSTKDVTVTDKNKI 153  
DB 1210 KLEEE--ADPTVSNLHQOVODRNEV 1233

RESULT 13  
ABG16636  
ID ABG16636 standard; Protein; 2519 AA.  
XX AC ABG16636;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #16627.  
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HSE-) HSEQ INC.  
XX PI Drmanac RT, Liu C, Tang YT;  
XX DR WPI: 2001-639362/73.  
XX DR N-PSDB; AAS80823.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX PT diagnostics, forensics, gene mapping, identification of mutations  
XX PT responsible for genetic disorders or other traits and to assess  
XX PT biodiversity  
XX PS Claim 20; SEQ ID NO 46995; 103bp; English.  
XX CC The invention relates to isolated polynucleotide (I) and  
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX CC and gene mapping, and in recombinant production of (II). The  
XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
XX CC to restore normal activity of (II) or to treat disease states involving  
XX CC (II). (II) is useful for generating antibodies against it, detecting or  
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
XX CC a food supplement. (II) and its binding partners are useful in medical  
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
XX CC disorders involving aberrant protein expression or biological activity.  
XX CC The polypeptide and polynucleotide sequences have applications in  
XX CC diagnostics, forensics, gene mapping, identification of mutations  
XX CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 2519 AA;  
QY Query Match 11.5%; Score 97.5; DB 22; Length 2519;  
Best Local Similarity 24.8%; Pred. No. 5.8;  
Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;  
QY 4 KEFLINKDTGSEVSELPKPHVTVTQNGKMSSTIVSEEDFILPYKGELEKGYOPDGEI 63  
DB 635 EKNWYKDKKPVVTEKTPSVTEKEVSKKPS-----PV-KAEVA-----EK 674  
QY 64 SGFEGKDAQVYINLSKDTFIKPVFKKIEKKKEENKPTFDVSKKKDNPQVNSQLNESH 123  
DB 675 QATDVKPPAAKKEKTVKKEKTPK-----EDKKEKEKPKKEVAKKEDKTPI---KKEKP 726  
QY 124 RKEDLQRE-----EHSQKSDSTKDV 143  
DB 727 KKEVKEKVKKEIKKEKKEKPEKEV 751

RESULT 14  
AAG37134  
ID AAG37134 standard; Protein; 408 AA.  
XX AC AAG37134;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 45610.  
XX KM Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.  
XX PR 09-MAR-1999; 99US-0123548.  
XX PR 23-MAR-1999; 99US-0125788.  
XX PR 25-MAR-1999; 99US-0126264.  
XX PR 29-MAR-1999; 99US-0126785.  
XX PR 01-APR-1999; 99US-0127462.  
XX PR 06-APR-1999; 99US-0128234.  
XX PR 08-APR-1999; 99US-0128714.  
XX PR 16-APR-1999; 99US-0129845.  
XX PR 19-APR-1999; 99US-0130077.  
XX PR 21-APR-1999; 99US-0130449.  
XX PR 23-APR-1999; 99US-0130510.  
XX PR 23-APR-1999; 99US-0130891.  
XX PR 28-APR-1999; 99US-0131449.  
XX PR 30-APR-1999; 99US-0132048.  
XX PR 30-APR-1999; 99US-0132407.  
XX PR 04-MAY-1999; 99US-0132484.  
XX PR 05-MAY-1999; 99US-0132485.  
XX PR 06-MAY-1999; 99US-0132486.  
XX PR 06-MAY-1999; 99US-0132487.  
XX PR 07-MAY-1999; 99US-0132863.  
XX PR 11-MAY-1999; 99US-0134256.  
XX PR 14-MAY-1999; 99US-0134218.  
XX PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 16-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143297.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159320.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.

PR	14-MAY-1999	99US-0134421
PR	14-MAY-1999	99US-0134370
PR	18-MAY-1999	99US-0134768
PR	19-MAY-1999	99US-0134941
PR	20-MAY-1999	99US-0135154
PR	21-MAY-1999	99US-0135523
PR	24-MAY-1999	99US-0135629
PR	25-MAY-1999	99US-0136021
PR	27-MAY-1999	99US-0136392
PR	28-MAY-1999	99US-0136782
PR	01-JUN-1999	99US-0137222
PR	03-JUN-1999	99US-0137528
PR	04-JUN-1999	99US-0137702
PR	07-JUN-1999	99US-0137724
PR	08-JUN-1999	99US-0138094
PR	10-JUN-1999	99US-0138540
PR	10-JUN-1999	99US-0138847
PR	14-JUN-1999	99US-0139119
PR	16-JUN-1999	99US-0139452
PR	16-JUN-1999	99US-0139453
PR	17-JUN-1999	99US-0139492
PR	18-JUN-1999	99US-0139454
PR	18-JUN-1999	99US-0139455
PR	18-JUN-1999	99US-0139460
PR	18-JUN-1999	99US-0139461
PR	18-JUN-1999	99US-0139463
PR	18-JUN-1999	99US-0139464
PR	18-JUN-1999	99US-0139750
PR	18-JUN-1999	99US-0139763
PR	21-JUN-1999	99US-0139817
PR	22-JUN-1999	99US-0139869
PR	23-JUN-1999	99US-0140353
PR	23-JUN-1999	99US-0140354
PR	24-JUN-1999	99US-0140695
PR	26-JUN-1999	99US-0140823
PR	29-JUN-1999	99US-0140931
PR	30-JUN-1999	99US-0141287
PR	01-JUL-1999	99US-0141841
PR	01-JUL-1999	99US-0142154
PR	02-JUL-1999	99US-0142055
PR	06-JUL-1999	99US-0142390
PR	08-JUL-1999	99US-0142803
PR	09-JUL-1999	99US-0142920
PR	12-JUL-1999	99US-0142977
PR	13-JUL-1999	99US-0143524
PR	14-JUL-1999	99US-0143624
PR	15-JUL-1999	99US-0144005
PR	16-JUL-1999	99US-0144085
PR	16-JUL-1999	99US-0144086
PR	19-JUL-1999	99US-0144321
PR	19-JUL-1999	99US-0144325
PR	19-JUL-1999	99US-0144332
PR	19-JUL-1999	99US-0144333
PR	19-JUL-1999	99US-0144343
PR	19-JUL-1999	99US-0144345
PR	19-JUL-1999	99US-0144352
PR	20-JUL-1999	99US-0144631
PR	20-JUL-1999	99US-0144684
PR	21-JUL-1999	99US-0144814
PR	21-JUL-1999	99US-0145086
PR	21-JUL-1999	99US-0145088
PR	22-JUL-1999	99US-0145085
PR	22-JUL-1999	99US-0145087
PR	22-JUL-1999	99US-0145089
PR	22-JUL-1999	99US-0145192
PR	23-JUL-1999	99US-0145145
PR	23-JUL-1999	99US-0145518
PR	23-JUL-1999	99US-0145524



PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 11.4%; Score 96.5; DB 21; Length 456;  
Best Local Similarity 25.0%; Pred. No. 0.67;  
Matches 40; Conservative 29; Mismatches 60; Indels 31; Gaps 8;

QY 10 KDTGEVSELRK-HRTVITQNGKMSSTVSEDFILPYKGELEKGYQFDGMEISGFEG 68  
DB 245 KKKEKDPKPKHPVSAFLVYANERRPALRENNKSVVEVAK-----ITGEKKNLSD 296  
QY 69 KKDAGY--VINLSDTFIKPVFKIEKKEENKPTFDVSKKKKDPQVNHSQLNESHRKE 126  
DB 297 KKKAPYEKXAKKKKETYLD-AMEBYKRTKEE-----ALSQKKE-----EEELKLKHGE 345  
QY 127 DLQREESHQKSDSTKDVTAFTVDKNNISKSTTN--PNK 164  
DB 346 ALQMLKKKEKTDN-----LIKKEKATKKKKENVDPRK 378

## RESULT 16

AAG37132  
ID AAG37132 standard; Protein; 476 AA.

XX AAG37132;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 45608.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KX termination sequence.

XX Arabidopsis thaliana.

OS EPI033405-A2.

PN EPI033405-A2.

XX 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138647.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140891.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145912.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147309.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 11.4%; Score 96.5; DB 21; Length 476;  
 Best Local Similarity 25.0%; Pred. No. 0.71;  
 Matches 40; Conservative 29; Mismatches 60; Indels 31; Gaps 8;

QY 10 KDTGEVSEIKP-HRVTTIQTQNGKMSSTIVSEDFILPYKGELEKGYOPDGMETSGFEG 68  
 DB 265 KNKKEKPLKPKPVSFAFLVYANERBALREKKSIVTEYAK-----ITGEEMNLSD 316  
 QY 69 KKDAGI--VINLSKDTFFIKVPFKKIEKEEKPFVDVSKKDNPOVNHSQLNESHKE 126  
 DB 317 KKKAPKPKYAKKKKTYLQ-AMEEYKRTKEE-----ALSQKE-----EEELLLKLKQE 365  
 QY 127 DLQREHSOKSDSTKDVATVLDKNNISSKSTNN--PNK 164  
 DB 366 ALQMLKKKEKTDN-----LKKKKATKKKKKNENVDPNK 398

RESULT 17  
 AAE20967  
 ID AAE20967 standard; Protein; 2060 AA.

AC AAE20967;  
 DT 01-JUL-2002 (first entry)  
 DE Staphylococcus lugdunensis von Willebrand factor binding protein.  
 KW Von Willebrand factor binding protein; vwb; immunogen; antibacterial;  
 OS Staphylococcus lugdunensis.  
 FH Key Location/Qualifiers  
 FT Peptide 1..47  
 FT Protein /label= Signal\_peptide  
 FT 48..2060  
 FT /note= "Mature von Willebrand factor binding protein"  
 PN MO200228892-A1.  
 PD 11-APR-2002.  
 PF 06-APR-2001; 2001MO-SE00766.  
 PR 04-OCT-2000; 2000SE-0003573.  
 PA (BIOS-) B10STAPRO AB.  
 PI Guss B, Frykberg L, Jacobsson K, Ahlen J, Nilsson M;  
 DR WPI; 2002-304928/34.  
 DR N-PSDB; AAD3371.  
 PT New von Willebrand factor binding protein from Staphylococci, useful  
 PT for determining and treating staphylococcal infection -  
 PS Claim 3; Page 35-41; 53pp; English.  
 CC The present invention relates to von Willebrand factor binding protein or  
 CC polypeptide (vwb) from Staphylococci. The vwb and immunogens of vwb are  
 CC useful in vaccines to combat infections caused by Staphylococci. The

CC invention is also useful for detection of staphylococcal infection and  
 CC purifying von Willebrand factor from a complex solution. The present  
 CC sequence is Staphylococcus lugdunensis vwb1 protein.

SQ Sequence 2060 AA;

Query Match 11.4%; Score 96.5; DB 23; Length 2060;  
 Best Local Similarity 25.7%; Pred. No. 5.5;  
 Matches 39; Conservative 25; Mismatches 49; Indels 39; Gaps 9;

QY 14 EVSEIKPRVYVITQ-NGKMSSTIVSEDFILPYKGELEKGYOPDG--WISGFEKXK 70  
 DB 1914 DIDEIK--ITTYVDNIGRE-----IVPSRKQPLP-EDFIGDMQYTHK-- 1955  
 QY 71 DAGVIYNLSKDTFFIKVPFKKIE-----KKEENKPFVDVSKKDNPOVNHSQLNESH 123  
 DB 1956 -----IKDGIITTYIKKVNANVPAKOLKTKTN--TOSBSQKHPQVQOLVKYHN 2006  
 QY 124 RKE--DLQREHSOKSDSTKDVATVLDKNNI 153  
 DB 2007 VKEQRSIEKSEHTDHSVSELPETGETANNGL 2038

RESULT 18  
 AAB18272  
 ID AAB18272 standard; Protein; 2500 AA.

AC AAB18272;  
 DT 07-NOV-2000 (first entry)  
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:129.  
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;  
 KW antimalarial; malaria; protozoacide; infection; insecticide.  
 OS Plasmodium falciparum.  
 PN MO200025728-A2.  
 PD 11-MAY-2000.  
 PF 05-NOV-1999; 99WO-US26796.  
 PR 05-NOV-1998; 98US-0107131.  
 PA (HOFF/) HOFFMAN S.  
 PA (CARU/) CARUCCI D.  
 PA (GARD/) GARDNER M.  
 PA (VENT/) VENTER J C.  
 PI Hoffman S, Carucci D, Gardner M, Venter JC;  
 DR WPI; 2000-365347/31.  
 PT Proteins encoded by chromosome 2 of the human malarial parasite,  
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
 PT diagnosis of P.falciparum infection -  
 PS Disclosure; Page 302-309; 577pp; English.  
 CC The present invention describes proteins and their fragments (I) encoded  
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
 CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)  
 CC vaccines against P. falciparum infection comprising (I) or (II).  
 CC (I) and (II) are useful for the development of vaccines against  
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal  
 CC antibody raised to immunogens comprising the sequences of (I), are  
 CC useful in the detection of infection with P. falciparum. Furthermore,  
 CC (I) (especially when they are rifins or secreted or membrane proteins)  
 CC can aid the identification of drugs to treat or prevent P. falciparum  
 CC infection, or they can be used to identify drug resistance in  
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the

CC subsequent identification of proteins encoded by it will help to expand  
CC our understanding of parasite biology, a process hampered by the  
CC complexity of the parasitic life cycle. Parasite resistance to drugs and mosquito  
CC vaccine and drug development. Parasite resistance to drugs and mosquito  
CC resistance to insecticides have led to a resurgence of malaria in many  
CC parts of the world, and there is a pressing need for vaccines and new  
CC drugs. AA70078 to AA70287 and AAB18144 to AAB18352 represent nucleotide  
CC and protein sequences given in the present invention, but which are not  
CC specifically mentioned within the specification.

XX Sequence 2500 AA;

Query Match 11.3%; Score 96; DB 21; Length 2500;

Best Local Similarity 26.1%; Pred. No. 8.1; Mismatches 54; Indels 32; Gaps 7;

Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;  
19 KPHRVTVTTONGKMSSTIVSEDFILPYKGELEKGYQPDGMEISGEGKKAGYVNL 78  
2173 KPYKIT---ENKK-----NEGNEILKKYSIENEKNNYDKEQENICILDKDTQCNYNT 2223

79 -----SKDTFIKVFVKIIEKKKEENKPTFDVSKKDNPOVNSQLNE---SHKED 127

2224 KEKNLNDKKSPFNISKYLEEEEEKSDKRD---DKNDNTEKKNLNDKKSFPFNISKYK 2280

128 LOREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164

2281 LEEB---KSDDKRD-----DKKNDNTEKKNLNDK 2308

RESULT 19

ABJ19106 standard; Protein; 645 AA.

ABJ19106;

06-MAR-2003 (first entry)

Pathogen specific antigen related staphylococcal protein SEQ ID No 414.

Antibacterial; virulence; fungicide; protozoicide; cytostatic; anti-HIV;  
hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;  
auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
autoimmune disease; HIV; hepatitis.

Staphylococcus sp.

WO200259148-A2.

01-AUG-2002.

21-JAN-2002; 2002WO-EP00546.

26-JAN-2001; 2001AT-0000130.

(CISTEM) CISTEM BIOTECHNOLOGIES GMBH.

Weinke A, Nagy B, Von Ahnen U, Klade C, Henics T, Zauner W,  
Mihh B, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M,  
Tempelmeier B;

WPI, 2003-075410/07.

Identifying, isolating and producing hyperimmune serum-reactive  
antigens from a pathogen, for preparing vaccine or medicament for  
treating or preventing e.g. staphylococcal infections, comprises  
providing antibody preparation

Example 7; Page 220; 252pp; English.

The invention relates to a novel method for identifying, isolating and  
producing hyperimmune serum-reactive antigens from a pathogen, tumour,  
allergen, a tissue or host prone to auto-immunity, where the antigens  
are used in a vaccine, comprises providing antibody preparation from a

CC plasma pool of a type of animal, or individual sera with antibodies  
CC against the specific pathogen, tumour, allergen, tissue or host prone to  
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of  
CC the 62 sequences of 53-2261 amino acids fully defined in the  
CC specification, or their hyperimmune fragments are useful for the  
CC manufacture of a pharmaceutical preparation, particularly a vaccine  
CC against staphylococcal infections or colonisation against S. aureus or S.  
CC epidermidis. The preparation of antibodies is useful for the manufacture  
CC of a medicament for treating or preventing staphylococcal infections or  
CC colonisation against S. aureus or S. epidermidis. The antibody  
CC preparations may also be used for diagnostic and imaging purposes. Other  
CC conditions that can be treated include cancer, autoimmune diseases or  
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or  
CC protozoan pathogens. This sequence represents a staphylococcal protein  
CC relating to the method for identifying and producing pathogen specific  
CC antigens of the invention.

XX Sequence 645 AA;

Query Match 11.0%; Score 93.5; DB 24; Length 645;

Best Local Similarity 21.9%; Pred. No. 2.2; Mismatches 74; Indels 67; Gaps 11;

Matches 49; Conservative 34; Mismatches 74; Indels 67; Gaps 11;  
1 TTVKEFLNKDTGCR-VSELEKPHRYTV--TTONGKMSSTIVSEDFILPYKGELE-KGY 56  
327 SAITEFQNVQPTNKKMDLQDTKYVYESVENNSMDTFYKH-----PIYGMINGKCY 381

57 QF-----DGWEISGFBGK-----D 71

382 MWMETTNDDYWKDPMVEGQVRITISKAKNNTRTIIPYVEGKTLYDAIVKHYKTIYD 441

72 AGVYINL-SKDTFIKVFVKIIEKKKEENKPTFDV---SKKDNPOVNSQLNESHRK 125

442 GQYHVRIVDKKAFPTKANTDSKKEQDQNSAKKATPATPSKPTPSVPEKESQKDSQKD 501

126 EDLQ---REHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164

502 DNKQPSVKEKENDASSBSGDKTPTKTKGVEVSSST--PTK 543

RESULT 20

ABP56879 standard; Protein; 654 AA.

ABP56879;

07-APR-2003 (first entry)

Staphylococcus epidermidis KrtN protein SEQ ID NO.10.

Staphylococcus aureus; Staphylococcus epidermidis; MSCRAMM; antibody;  
microbial surface component recognising adhesive matrix molecule;  
surface protein; infection; antibacterial; antiinflammatory; vaccine;  
immunosuppressive; antiallergic; gene therapy; pneumonia; endocarditis;  
septic arthritis; biomaterial related infection.

Staphylococcus epidermidis.

WO2002102829-A2.

27-DEC-2002.

17-JUN-2002; 2002WO-US19220.

15-JUN-2001; 2001US-298098P.

(INHIBITEX INC. (QUBES) QUBEN ELIZABETH COLLEGE DUBLIN.

(UYPA-) UNIV PAVIA.

Foster TJ, Roche F, Patti JM, Hutchins JT, Hall A, Domanski P;  
Patel P, Stribeys P, Speziale P;



AC AAG82397;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis open reading frame protein sequence SEQ ID NO:1898.  
XX  
KM Staphylococcus epidermidis SRI strain; infection; diagnosis;  
XX vaccination; endocarditis.  
XX  
OS Staphylococcus epidermidis.  
XX  
FN MO200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000MO-US30782.  
XX  
PR 09-NOV-1999; 99US-0164258.  
XX  
PA (GLAX ) GLAXO GROUP LTD.  
XX  
PI Kimmery MJ;  
XX  
DR WPI; 2001-316495/33.  
DR N-PSDB; AAH53247.  
XX  
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
XX  
PS Claim 18; Page 516; 2188bp; English.  
XX  
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce host cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.  
XX  
SQ Sequence 299 AA;  
Query Match 10.8%; Score 92; DB 22; Length 299;  
Best Local Similarity 24.4%; Pred. No. 1;  
Matches 39; Conservative 27; Mismatches 60; Indels 34; Gaps 5;  
QY 9 NKDTGEVSELKPHRTVTYIIONGKEMSTIVSEEDFI---LPVYKGELEKG-YOPDGWEI 63  
DB 33 NKDT-EKSDKTKYHRIISLPSTELVRLGIGEDIVGSTVDYDKVKKGQKDANML 91  
QY 64 SGFE-----GKKDAGVYINLSKDTFIKPY-----FKKIEKK 95  
DB 92 NKEELIKAKPLLILAHESQKNSAGKVLKSLDKGVKVVYVVDQAQSIDETDTPFSIGOLT 151  
QY 96 EENKPTFDVSKKKDNPQVNSQLNESHKEDLQREHSQ 135  
DB 152 DREKQAKELVDETGNVEKTIINSVPKHHKQEVMEVSSK 191

ID ABP40822 standard; Protein; 309 AA.  
XX  
AC ABP40822;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5667.  
XX  
KM Staphylococcus epidermidis, open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
FN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI; 2002-381255/41.  
DR N-PSDB; ABN93367.  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 5667; 267bp; English.  
XX  
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP5124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 309 AA;  
Query Match 10.8%; Score 92; DB 23; Length 309;  
Best Local Similarity 24.4%; Pred. No. 1.1;  
Matches 39; Conservative 27; Mismatches 60; Indels 34; Gaps 5;  
QY 9 NKDTGEVSELKPHRTVTYIIONGKEMSTIVSEEDFI---LPVYKGELEKG-YOPDGWEI 63  
DB 43 NKDT-EKSDKTKYHRIISLPSTELVRLGIGEDIVGSTVDYDKVKKGQKDANML 101  
QY 64 SGFE-----GKKDAGVYINLSKDTFIKPY-----FKKIEKK 95  
DB 102 NKEELIKAKPLLILAHESQKNSAGKVLKSLDKGVKVVYVVDQAQSIDETDTPFSIGOLT 161  
QY 96 EENKPTFDVSKKKDNPQVNSQLNESHKEDLQREHSQ 135  
DB 162 DREKQAKELVDETGNVEKTIINSVPKHHKQEVMEVSSK 201

RESULT 24  
ABP38188  
ID ABP38188 standard; Protein; 442 AA.  
XX  
AC ABP38188;  
XX  
DT 24-JUL-2002 (first entry)  
XX

DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3033.  
XX  
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
KM antibacterial; gene therapy.  
XX  
XX Staphylococcus epidermidis.  
OS  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
PI  
DR WPI; 2002-381255/41.  
DR N-PSDB; ABN90733.  
XX  
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
PS  
XX Disclosure; SEQ ID 3033; 267pp; English.  
XX  
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 442 AA;  
XX  
Query Match 10.7%; Score 91; DB 23; Length 442;  
Best Local Similarity 21.7%; Pred. No. 2.2;  
Matches 49; Conservative 30; Mismatches 65; Indels 82; Gaps 11;  
QY 10 KDTGEVSELPKPHVTVT-IQNGKE--MSSTIVSEEDFILPVYKG-ELEK----- 54  
DB 90 EDT-EINPVAQPEVNTQIEKGDFEATVVEPEVKLGADYKGLIEKEQETDLSDEBQ 148  
QY 55 -----GVQFDGMEISG--FECKKAGVYINISKDTFTK 85  
DB 149 ESIDHSLSLAEVNVKEDGAVNGDTVINIDFSG-SVDGEFFDGGQAGEYDLETGSSGFLP 207  
QY 86 PVFKKIEEKKKEENK-----PTFDVS---KKKDNPOVNHSLQNE- 121  
DB 208 GFEEQIEGKTKGDEKVVVTFPEEYHAEELAGKEATFKKVNKIKKQVPELNDIELANEL 267  
QY 122 -----SHRKEDLQREHHSQKSDSTKDVATVATVLDKNINISKSSTNN 160  
DB 268 DSDAENVDEYKENLRRKRLSEQKATEAMNT-----EKKEALINKATEN 308  
XX  
RESULT 25  
ID AAB40312  
XX AAB40312 standard; Protein; 902 AA.  
XX  
XX AAB40312;  
XX  
XX 24-JUL-2002 (first entry)  
XX  
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:5157.  
XX

KM Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.  
XX  
XX Staphylococcus epidermidis.  
OS  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
PI  
DR WPI; 2002-381255/41.  
DR N-PSDB; ABN92857.  
XX  
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
PS  
XX Disclosure; SEQ ID 5157; 267pp; English.  
XX  
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABB35124 to ABB37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compounds able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 902 AA;  
XX  
Query Match 10.7%; Score 91; DB 23; Length 902;  
Best Local Similarity 24.9%; Pred. No. 6.1;  
Matches 43; Conservative 31; Mismatches 75; Indels 24; Gaps 8;  
QY 1 TTYKPEILNKDGEVSELPKPHVTVTQNGKMS-STIVSEEDFILPVYKGLKGYQPD 59  
DB 736 TPVSEYRISNRGKGKIK---TATYTERGNITCITTVGEBDLNVVTVNAGVI---IRLD 788  
QY 60 GWEISGFESGKXAGY-VINISKDTFLPVFKKIEEKKKEEN-----KPTFDVSKKK 109  
DB 789 VHDISQ-NGRAAQGRIMKLGCGQGVSTVAKXNEEDDNEENDEAQSSTTETADVEEV 847  
QY 110 DNPQVNHSLNESHKREDIQREE-HSQKSDSTKDVATVATVLDKNINISKSSTNN 161  
DB 848 D-----DQTPGNAIHTEGDAEMESVESPENDDRIDIRODFMDRVNEDIBSASDN 896  
XX  
RESULT 26  
ID AAB42897  
XX AAB42897 standard; Protein; 209 AA.  
XX  
XX AAB42897;  
XX  
XX 08-FEB-2001 (first entry)  
XX  
XX Human ORFX ORF2661 polypeptide sequence SEQ ID NO:5322.  
XX  
XX  
XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;  
XX vulnerable; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;  
XX anticonvulsant; osteopachytic; antiarthritic; immunosuppressant; cardiant;  
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;  
XX antiviral; antibacterial; antifungal; antineumatic; antichyroid;  
XX



XX	antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus erythematosus; infection;
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW	thrombosis; contraceptive.
XX	
XX	Homo sapiens.
OS	
XX	
PN	WO200058473-A2.
XX	
XX	
PD	05-OCT-2000.
XX	
PF	31-MAR-2000; 2000WO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.
PR	02-APR-1999; 99US-0127636.
PR	05-APR-1999; 99US-0127728.
PR	30-MAR-2000; 2000US-0540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shimkete RA, Leach M;
XX	
DR	WPI; 2000-602362/57.
DR	N-PSDB; AAC77106.
XX	
XX	
PT	Novel nucleic acids and peptides derived from open reading frame X,
PT	useful for treating e.g. cancers, proliferative disorders,
PT	neurodegenerative disorders and cardiovascular disease -
XX	
XX	
XX	Claim 11; Page 4497-4498; 5507jp;English.

Query Match	10.7%	Score 90.5	DB 21	Length 209
Best Local Similarity	28.1%	Pred. No. 0.89		
Matches	47	Conservative 17	Mismatches 56	Indels 47
				Gaps 9
QY	13	GEVSLKPRVVTIYONGKEMSTIYSE--EDILTPVYGELEKGFQ-----DGM-----	62	
Db	18	GLIQLSGSL-----HGKSDSPNVYTEKEKIALIERLTLEKLTFFQQRSDLWERLY	72	
QY	63	-----ISFGKKDAG-----YVNIISKDTPIKVFPKKIEKKENKPFEDVSKKD	110	
Db	73	VEAKDQNGKQGTGDKKKGGSHRANKSKETFFLSV-----KETFDAMKNST	120	
QY	111	NQGVNH-----SQMNSHRKEDLQREHRSQSD--STKDVATATVDK	150	

DB 121 KEFVHHKIKOAKEA-VKENLKKFSDSVKSTFRHFKDTKNIPDE 166

RESULT 27  
ID ABP73209 standard; Protein; 635 AA.  
XX ABP73209;  
AC  
XX  
DT 30-JAN-2003 (first entry)  
XX  
DE Candida albicans essential protein SEQ ID NO 7046.  
XX  
XX Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;  
KW signal transduction; DNA replication; cell division; growth;  
KM proliferation; Candida albicans; fungicide; antifungal.  
XX  
OS Candida albicans.  
XX  
FN WO200253728-A2.  
XX  
PD 11-JUL-2002.  
XX  
PF 26-DEC-2001; 2001WO-US49486.  
XX  
PR 29-DEC-2000; 2000US-259128P.  
XX 20-FEB-2001; 2001US-079202A.  
PR 22-AUG-2001; 2001US-314050P.  
XX  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
XX  
XX WPI; 2002-566694/60.  
DR N-PSDB; ABZ31759.  
XX  
XX Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele  
PT of a gene and placing other allele of the gene under conditional  
PT expression -  
XX  
XX Claim 44; SEQ ID NO 7046; 167pp + Sequence listing; English.

Query Match	10.7%;	Score 90.5;	DB 23;	Length 635;
Best Local Similarity	22.2%;	Pred. No. 4.2;		

Matches 38; Conservative 40; Mismatches 70; Indels 23; Gaps 8;

QY 6 FILKDTGEVSELSKPH---RVVTIONGKEM-----SSRIVEEEDFILPVYGELEKGY 56  
 Db 64 YTMNKE--KKKKFKSPPTMTPTPKOGILNHWKSGKELLIOQE---EEDGTPKPY 117  
 QY 57 QPDGM---EISGF--EGKKDAGVINTLSK-DTFIKPVFKLEKEKEENKPTFDVSKKK 109  
 Db 118 FFEKYNVQLELPFVEDPTDYL-YMIEIKYESKMEERARRKEREKREKRDLEKKKK 176  
 QY 110 DNPQVHSGQLNESHKREDLQREHSGKSOSTKVYATVLDKNISKSTTN 160  
 Db 177 QQQQQQKSOONPQNIKDKDEKNQDTRNNTDKDSEKSEDKPTVEAKKED 227

RESULT 28  
 AA44364  
 ID AA44364 standard; protein; 757 AA.  
 AC AA44364;  
 XX  
 XX 14-MAR-2000 (first entry)  
 DE Human cell cycle regulation protein-5.  
 XX  
 KM CECRP-5; cell cycle regulation protein-5; cell proliferation;  
 KM cell proliferative disease; cancer; atherosclerosis; cirrhosis;  
 KM hepatitis; psoriasis; immune system disorder; allergy; asthma;  
 KM acquired immune deficiency syndrome; Crohn's disease; Blast method;  
 KM rheumatoid arthritis; gene therapy; chromosomal mapping.  
 XX  
 OS Homo sapiens.  
 XX

Key Location/Qualifiers  
 FT Modified-site 4 /note= "Potential glycosylation site"  
 FT Modified-site 63 /note= "Potential glycosylation site"  
 FT Modified-site 479 /note= "Potential glycosylation site"  
 FT Modified-site 498 /note= "Potential glycosylation site"  
 FT Modified-site 31 /note= "Potential glycosylation site"  
 FT Modified-site 43 /note= "Potential phosphorylation site"  
 FT Modified-site 81 /note= "Potential phosphorylation site"  
 FT Modified-site 96 /note= "Potential phosphorylation site"  
 FT Modified-site 99 /note= "Potential phosphorylation site"  
 FT Modified-site 130 /note= "Potential phosphorylation site"  
 FT Modified-site 158 /note= "Potential phosphorylation site"  
 FT Modified-site 162 /note= "Potential phosphorylation site"  
 FT Modified-site 163 /note= "Potential phosphorylation site"  
 FT Modified-site 184 /note= "Potential phosphorylation site"  
 FT Modified-site 186 /note= "Potential phosphorylation site"  
 FT Modified-site 188 /note= "Potential phosphorylation site"  
 FT Modified-site 271 /note= "Potential phosphorylation site"  
 FT Modified-site 277 /note= "Potential phosphorylation site"  
 FT Modified-site 287 /note= "Potential phosphorylation site"  
 FT Modified-site 289 /note= "Potential phosphorylation site"

FT Modified-site /note= "Potential phosphorylation site"  
 FT Modified-site 323 /note= "Potential phosphorylation site"  
 FT Modified-site 343 /note= "Potential phosphorylation site"  
 FT Modified-site 367 /note= "Potential phosphorylation site"  
 FT Modified-site 370 /note= "Potential phosphorylation site"  
 FT Modified-site 432 /note= "Potential phosphorylation site"  
 FT Modified-site 488 /note= "Potential phosphorylation site"  
 FT Modified-site 499 /note= "Potential phosphorylation site"  
 FT Modified-site 207 /note= "Potential phosphorylation site"  
 FT Modified-site 277 /note= "Potential phosphorylation site"  
 FT Modified-site 298 /note= "Potential phosphorylation site"  
 FT Modified-site 370 /note= "Potential phosphorylation site"  
 FT Modified-site 419 /note= "Potential phosphorylation site"  
 FT Modified-site 474 /note= "Potential phosphorylation site"  
 FT Modified-site 488 /note= "Potential phosphorylation site"  
 FT Modified-site 499 /note= "Potential phosphorylation site"  
 FT Modified-site 528 /note= "Potential phosphorylation site"  
 FT Modified-site 532 /note= "Potential phosphorylation site"  
 FT Modified-site 539 /note= "Potential phosphorylation site"  
 FT Modified-site 609 /note= "Potential phosphorylation site"  
 FT Modified-site 624 /note= "Potential phosphorylation site"  
 FT Modified-site 730 /note= "Potential phosphorylation site"  
 FT Modified-site /note= "Potential phosphorylation site"  
 PN MO9964593-A2.  
 XX 16-DEC-1999.  
 PD 08-JUN-1999; 99WO-US12906.  
 PF 08-JUN-1998; 98US-0088695.  
 PR (INCY-) INCYTE PHARM INC.  
 PA Bandman O, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR,  
 PI Paterson C;  
 PI WPI; 2000-105867/09.  
 DR N-PSDB; AA229484.  
 XX Novel regulatory proteins, for diagnosis, treatment and prevention of  
 PT cell proliferative and immune system diseases  
 XX  
 XX Claim 1; Page 70-71; 87pp; English.  
 PS  
 XX  
 CC The present sequence is cell cycle regulation protein-5 (CECRP-5). Blast  
 CC analytical method was used to identify this protein. CECRPs are  
 CC activators of cell proliferation or inhibitors of cellular processes that  
 CC modulate proliferation. They are used to treat or prevent cell  
 CC proliferative diseases like cancers, atherosclerosis, cirrhosis,  
 CC hepatitis, psoriasis, immune system disorders (e.g. acquired immune  
 CC deficiency syndrome, allergy, asthma, Crohn's disease, rheumatoid  
 CC arthritis). Antibodies are raised to screen for specific binding agents.

CC The corresponding nucleic acid is used in gene therapy, chromosomal  
CC mapping and isolation of related sequences.

XX Sequence 757 AA;

Query Match 10.7%; Score 90.5; DB 21; Length 757;

Best Local Similarity 28.1%; Pred. No. 5.3;

Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 13 GEVSELKPHRYVTITONGKEMSTIVSE--EDFILPVYKGELEKGYQF-----DGMF--- 62

DB 398 GELQOLSGSSQI-----HGKSDSPNVYTEKEIALIRERLTLEKLTFTQQRSDIMERLY 452

QY 63 -----ISGFEKKDAG---YVINLSKDTFIPVKKLEKKEENKPTFVSKKCD 110

DB 453 VEAKDQNGKQGTGDKKKGGSHRAKKSKETFLGSV-----KETFDAMKNST 500

QY 111 NPQVNH-----SQLNESHKEDLQREHSQKSD--STKDVATVLDK 150

DB 501 KEFVRHHKEKIKQAKEA-VKENLKKFSDSVKSTFRHFDTTKNIFDE 546

#### RESULT 29

AAB53319

ID AAB53319 standard; Protein; 758 AA.

XX AAB53319;

DT 09-MAR-2001 (first entry)

XX Human colon cancer antigen protein sequence SEQ ID NO:859.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW identification; cytostatic; cardioprotective; neuroprotective; vulnary;  
KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
KW nephrologic; antiinfective; antibacterial; gene therapy; wound;  
KW neutral disorder; immune system disorder; muscular disorder;  
KW reproductive disorder; gastrointestinal disorder; renal disorder;  
KW infectious disease; cardiovascular disorder.

XX Homo sapiens.

XX WO200055351-A1.

XX 21-SEP-2000.

XX 08-MAR-2000; 2000WO-US05883.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2000-587534/55.

XX N-PSDB; AAC98076.

XX Colon cancer associated gene sequences, referred to as colon cancer  
PT antigens, useful for the treatment, prevention, and diagnosis of colon  
XX disorders such as colon cancer -

XX Claim 11; Page 1408-1411; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
CC called human colon cancer antigens, given in AAB53319 to AAB54006. The  
CC human colon cancer antigens can have cytostatic, cardioprotective, muscular;  
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
CC vulnary, nephrologic, antiinfective and antibacterial activities, and  
CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
CC proteins and antibodies to the proteins are useful for the prevention,  
CC treatment and diagnosis of colon disorders, such as colon cancer. The  
CC polynucleotides may be used in diagnostics and research, such as for  
CC chromosome identification, and as hybridisation probes. The proteins

CC may also be used to prevent diseases such as neural disorders, immune  
CC system disorders, muscular disorders, reproductive disorders,  
CC gastrointestinal disorders, wounds, renal disorders, infectious  
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
CC AAB54007 represent sequences used in the exemplification of the present  
CC invention.

XX Sequence 758 AA;

Query Match 10.7%; Score 90.5; DB 21; Length 758;

Best Local Similarity 28.1%; Pred. No. 5.3;

Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;

QY 13 GEVSELKPHRYVTITONGKEMSTIVSE--EDFILPVYKGELEKGYQF-----DGMF--- 62

DB 399 GELQOLSGSSQI-----HGKSDSPNVYTEKEIALIRERLTLEKLTFTQQRSDIMERLY 453

QY 63 -----ISGFEKKDAG---YVINLSKDTFIPVKKLEKKEENKPTFVSKKCD 110

DB 454 VEAKDQNGKQGTGDKKKGGSHRAKKSKETFLGSV-----KETFDAMKNST 501

QY 111 NPQVNH-----SQLNESHKEDLQREHSQKSD--STKDVATVLDK 150

DB 502 KEFVRHHKEKIKQAKEA-VKENLKKFSDSVKSTFRHFDTTKNIFDE 547

#### RESULT 30

ABP68963

ID ABP68963 standard; Protein; 801 AA.

XX ABP68963;

DT 20-JAN-2003 (first entry)

XX Human polypeptide SEQ ID NO 1010.

XX Human; genome mapping; gene therapy; food supplement; virus; fungus;  
KW cell-proliferative disorder; neurodegenerative disease; bacterial;  
KW Parkinson's disease; Alzheimer's disease; autoimmune disease;  
KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
KW arthritis; cytostatic; immunomodulator; nocotropic; neuroprotective;  
KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;  
KW haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;  
KW antichrilitic.

XX Homo sapiens.

XX WO200270539-A2.

XX 12-SEP-2002.

XX 05-MAR-2002; 2002WO-US05095.

XX 05-MAR-2001; 2001US-0799451.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;

XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;

XX Wehman T, Wang J, Wang D, Drmanac RT;

XX WPI; 2002-759812/82.

XX N-PSDB; AB211180.

XX New polynucleotides comprising sequences assembled from expressed  
PT sequence tags (ESTs), useful for treating cell-proliferative,  
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
PT platelet or coagulation disorders -

XX Claim 9; SEQ ID NO 1010; 1012pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (I) comprising a  
CC nucleotide sequence selected from any of 948 sequences

CC (ABZ1119-ABZ12066) or their mature protein coding portion, active domain  
CC coding protein or complementary sequences. The polynucleotides are useful  
CC for identifying expressed genes or for physical mapping of human genome.  
CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular  
CC weight markers, as a food supplement, for generating antibodies, in  
CC medical imaging, screening and diagnostic assays and for treating  
CC cell-proliferative disorders (cancer), neurodegenerative diseases  
CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid  
CC disorders, platelet or coagulation disorders, wound, burn, incision,  
CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
CC parasitic), arthritis, etc.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 801 AA;  
Query Match 10.7%; Score 90.5; DB 23; Length 801;  
Best Local Similarity 22.1%; Pred. No. 5.8; Mismatches 31; Gaps 6;  
Matches 34; Conservative 35; Indels 31; Gaps 6;  
QY 4 KEFLNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEDFILFVYKGELEKGYQPDGWEI 63  
DB 504 KEFLNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEDFILFVYKGELEKGYQPDGWEI 551  
QY 64 SCFEGKKDAGYVNLKDTFIKVFVKIEKKEENKPTFDV-----SKKKNPQVNHSQL 119  
DB 552 -----DLGOELRLTRERQ-VQNSHTELAERHQVQAOERIEIRLSLEDMKLSKEXL 603  
QY 120 NESHRKEDLQREHESOKSDSTKQVTAIVLTKNNI 153  
DB 604 -----EKAKLELEE---AQDTVSNLHQOVQDRNEV 630  
RESULT 31  
AAG85023  
ID AAG85023 standard; Protein; 922 AA.  
AC AAG85023;  
XX  
XX  
DT 11-SEP-2001 (first entry)  
XX  
DE Shrimp white spot Bacilliform virus (WSBV) protein 114.  
XX  
XX Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;  
KW antiviral agent; gene expression; antisense construct;  
KW transgenic viral resistant shrimp.  
XX  
OS White spot syndrome virus.  
XX  
PN WO200136351-A2.  
XX  
PD 31-MAY-2001.  
XX  
PF 08-NOV-2000; 2000WO-US28888.  
XX  
PR 24-NOV-1999; 99CN-0124717.  
XX  
PA (PENY-) PE CORP NY.  
PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.  
PA (SINO-) SINOGENOMAX CO LTD.  
XX  
XX Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C,  
XX  
XX WPI; 2001-355877/37.  
XX  
XX N-PSDB; AAH62803.  
PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus  
PT (WSBV), useful for producing viral polypeptides that can be used to  
PT screen for agents that are useful for treating WSBV infection -  
XX  
PS Claim 1; Figure 3; 626bp; English.

XX  
CC The invention provides the primary nucleotide sequence of the WSBV genome  
CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and  
CC encoded proteins (AAH64910-AAH65051) and oligonucleotide sequences  
CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid  
CC molecules and proteins of the invention are useful for diagnosis and  
CC monitoring viral infection, in screens for antiviral agents and for  
CC monitoring viral gene expression or activity during a treatment regimen.  
CC The nucleic acid molecules are also useful as antisense constructs to  
CC control viral gene expression in infected cells and tissues and to create  
CC transgenic viral resistant shrimp.  
XX  
SQ Sequence 922 AA;  
Query Match 10.7%; Score 90.5; DB 22; Length 922;  
Best Local Similarity 22.0%; Pred. No. 7; Mismatches 68; Indels 51; Gaps 8;  
Matches 42; Conservative 30; Indels 51; Gaps 8;  
QY 10 KDTGEVSELKPHRVTVTIQNG-----KMSSTIVSEDFILFVYKGELEKGYQPDGWEI 48  
DB 300 KTTNQIVE---ESVYESRNGCCCFKNDKWLAKRESNLKLNNTVFGEDDEKSAVAYSD 356  
QY 49 -----KGELEKGYQPDGWEISGFEKKDAGYVNLKDTFIKVF-FKIEKKEENK 100  
DB 357 SEDEDEDENEEVDYDYNNETIESVG-----NVIKNLIRKTIQGLDVEEKEEGEQ 408  
QY 101 PTFVDSKKKDNQVNHSQLNESHRKEDL-----QREHESOKSDSTKQVTAIVL-----DK 150  
DB 409 SEEEEDSDDDDDASVSSSSSSSVTVVAAAEEDDEDEKOKD-TATVDEDEDK 467  
QY 151 NNISKSTTN 161  
DB 468 ESVISSSSSDS 478  
RESULT 32  
ABB58769  
ID ABB58769 standard; Protein; 1183 AA.  
AC ABB58769;  
XX  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 3099.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
XX Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
XX N-PSDB; ABL02872.  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure; SEQ ID NO 3099; 21bp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB57737-AB872072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 1183 AA;

Query Match 10.6%; Score 90; DB 22; Length 1183;

Best Local Similarity 24.7%; Pred. No. 11; Mismatches 54; Indels 33; Gaps 5;

Matches 36; Conservative 24; Mismatches 54; Indels 33; Gaps 5;

QY 19 KPHRTVTITQNGKMSSTIVSEEDFLPYKGELEKGYQPDGMEISGFEKKDAGVIVL 78  
DB 310 KPEVTTAKDNKKD---PLIADSKPKVEPKETSK-----TGKKGKGA----- 352

QY 79 SKDTFKVFFKKIEKKKEENKPTDVSKKKONPOVNHSQLNESHKEDLQREHSQ--- 135  
DB 353 -----KPAEK--SKKEEKEDSTTKSKKKKADSPANNQKEQIGYKKTSEPDASNSHK 404

QY 136 ----KSDTKDVTATVLDKNISSKS 157  
DB 405 SVTAKKDSKDEASTQVKSNDNSPEA 430

RESULT 33

AAV03190  
ID AAV03190 standard; Protein; 402 AA.

XX AAV03190;

DT 16-JUN-1999 (first entry)

DE S. aureus trigger factor protein sequence.

XX Trigger factor; tig; diagnosis; immune response; bacterial infection;  
KW S. aureus infection; otitis media; toxic shock syndrome; conjunctivitis;  
KW wound infection; impetigo; septic arthritis; therapy.

OS Staphylococcus aureus.

XX EP905233-A2.

PD 31-MAR-1999.

PF 21-AUG-1998; 98EP-0306697.

PR 29-DEC-1997; 97US-0999339.

PR 04-SEP-1997; 97US-0057511.

PA (SMK ) SMITHKLINE BEECHAM CORP.

PA (SMK ) SMITHKLINE BEECHAM PLC.

XX Burnham MKR, Fosberry A, Hodgson JE, Jaworski DJ;

PI Lawlor EJ, Rosenberg M, Traini CM, Wang M, Ward J;

PI Warren RL;

DR WPI; 1999-192663/17.

DR N-PsDB; AAX28175.

XX New Staphylococcus aureus trigger factor (tig) polypeptide and

PT polynucleotide, useful as diagnostic reagents and for prevention and

PT treatment of Staphylococcus aureus infections

XX Claim 23; Page 8; 31pp; English.

XX This sequence is the Staphylococcus aureus trigger factor (tig)

CC polypeptide of the invention. Tig polypeptides and polynucleotides are

CC useful for diagnosing diseases related to over or underexpression of tig  
CC protein by identifying mutations in the tig gene, or determining tig  
CC polypeptide or mRNA expression levels due to an infection of an organism  
CC with the tig gene. They can diagnose the stage and type of infection. Tig  
CC polypeptides are also useful for screening compounds which affect  
CC activity of the protein by measuring the binding to tig and observing the  
CC stimulation or inhibition of the polypeptide function. These can be used  
CC in treatment to inhibit or enhance tig activity, in addition to direct  
CC administration of tig polypeptides to treat conditions associated with a  
CC lack of tig polypeptide, or direct administration of antisense sequences  
CC to prevent expression. Tig polypeptides (administered directly, in a  
CC vector and as a vaccine) and antibodies induce an immune response to  
CC immunise and prevent disease. Diseases diagnosed, prevented or treated  
CC include: bacterial infections, especially *Staphylococcus aureus*  
CC infections which cause otitis media, toxic shock syndrome,  
CC conjunctivitis, wound infection, impetigo and septic arthritis etc. Tig  
CC polypeptides, polynucleotides and their (ant)agonists can prevent  
CC adhesion of bacteria to matrix proteins, and are useful for use on wounds  
CC and body implants to prevent bacterial infection.

XX Sequence 402 AA;

Query Match 10.6%; Score 89.5; DB 20; Length 402;

Best Local Similarity 22.1%; Pred. No. 2.8; Mismatches 60; Indels 83; Gaps 13;

QY 11 DTGEVSELKPHRYVT-IONKE--MSSTIVSEEDFLPYKKG-FLER----- 54  
DB 53 DETDIKPYAOPESVTOIEKDPFEATVVEPVKIDGYKGLIEKQETELSDDELOE 112

QY 55 -----GYQPDGMEISG--FEGKKDAGVIVLSDKTFIKP 86  
DB 113 AIDHSLGHLAEVWVEDGVENGDTVINDFSG-SVDGSEFEGQAGEYDLRGSSSFI-P 170

QY 87 VFK-----KIEKKE-----EE--NKPTPDVS---KKONPOVNHSQLNE- 121  
DB 171 GFEEQLGKMKVDEKDVVTFPEEYHAELELAKGKATFTKVNVEIKFKEVPELTDIENEL 230

QY 122 -----SHRKEDLQREHSQSDSTKDYTVATVLDKNISSKSTTN 160  
DB 231 DAEANTVDEYKENLKRILAEQKATDAENV-----EKEBATYKATDN 271

RESULT 34

AAV35091  
ID AAV35091 standard; Protein; 511 AA.

XX AAV35091;

DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae transmembrane protein sequence.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

XX WO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

PA (GENSET ) GENSET.

XX Griffiths R;

XX WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae  
PT  
XX  
PS Page 975-976; Disclosure; 1912pp; English.  
XX  
CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
CC frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.  
CC C. pneumoniae causes respiratory disease such as pneumonia and  
CC bronchitis and is thought to be a contributing factor in heart  
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
CC nucleotide sequences can also be used as immunogenic compositions,  
CC especially where the vector directs the expression of a neutralising  
CC epitope of C. pneumoniae.  
XX  
SQ Sequence 511 AA;  
XX  
Query Match 10.6%; Score 89.5; DB 20; Length 511;  
Best Local Similarity 24.5%; Pred. No. 3.9;  
Matches 23; Conservative 17; Mismatches 33; Indels 21; Gaps 2;  
QY 84 IXPVFKIEBKEENKPTFD-----VSKKQDPQVNHSQLNESHAK 125  
DB 95 VKGVFKTPOAREVSPRLPSHVHGQRLPGLGFRDRIOKRSKSEPEADLGKMKRSYSD 154  
QY 126 EDIQREHSQKSDSTQVATVYIDKNNISKSTT 159  
DB 155 GDIDRVGHDSNEDSTEDSRS---EGGEPSKSSS 185  
XX  
RESULT 35  
ABP73570  
ID ABP73570 standard; Protein; 516 AA.  
XX  
AC ABP73570;  
XX  
DT 30-JUN-2003 (first entry)  
XX  
DE Candida albicans essential protein SEQ ID NO 7407.  
XX  
KM Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;  
KM signal transduction; DNA replication; cell division; growth;  
KM proliferation; Candida albicans; fungicide; antifungal.  
XX  
OS Candida albicans.  
XX  
OS  
XX  
PN WO200253728-A2.  
XX  
PD 11-JUL-2002.  
XX  
PF 26-DEC-2001; 2001WO-US49486.  
XX  
PR 29-DEC-2000; 2000US-259128P.  
PR 20-FEB-2001; 2001US-0792024.  
PR 22-AUG-2001; 2001US-314050P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;  
XX  
DR WFI, 2002-566694/60.  
DR N-PSDB; AB232120.  
XX  
PT Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele  
PT of a gene and placing other allele of the gene under conditional  
PT expression -  
XX  
PS Claim 44; SEQ ID NO 7407; 167bp + Sequence Listing; English.  
XX  
CC The invention relates to constructing (M1) a strain of diploid fungal

CC cells in which both alleles of a gene are modified, comprising modifying  
CC one allele by insertion or replacement by a cassette having an  
CC expressible selectable marker and modifying other allele by  
CC recombination of a promoter replacement fragment with a heterologous  
CC promoter, so that expression of the second allele is regulated by the  
CC promoter. (M1) is useful for constructing a strain of diploid fungal  
CC cells in which both alleles of a gene are modified. The diploid fungal  
CC cells having both alleles modified are useful for identifying a gene that  
CC is essential to the survival or growth of a fungus, a gene that  
CC contributes to the virulence and/or pathogenicity of a fungus, a gene  
CC that contributes to the resistance of a diploid fungus to an antifungal  
CC agent, an antifungal agent that inhibits the growth of a diploid fungus  
CC and for identifying a therapeutic agent for treatment of a mammalian  
CC disease. (M1) is useful for identifying a compound which modulates the  
CC activity of a gene product, preferably enzymatic activity, carbon  
CC compound catabolism, biosynthetic, transporter, transcriptional,  
CC translational, signal transduction, DNA replication and cell division  
CC activity. The method is useful for identifying a compound having the  
CC ability to inhibit growth or proliferation of C. albicans cells and for  
CC treating infection by C. albicans. The present sequence is that of an  
CC essential Candida albicans protein used in the method of the invention.  
CC Note: The sequence data for this patent is not represented in the printed  
CC specification but is based on sequence information supplied to Derwent by  
CC the European Patent Office.  
XX  
SQ Sequence 516 AA;  
XX  
Query Match 10.6%; Score 89.5; DB 23; Length 516;  
Best Local Similarity 23.3%; Pred. No. 3.9;  
Matches 37; Conservative 27; Mismatches 58; Indels 37; Gaps 6;  
QY 10 KDTGEVSELKPRVTVT--IQNGKEMSPRIVEEDFILPVYKGELEKGYQF----- 58  
DB 369 KVKGRISRYLAKKCISRIDVYSEPTTAFGE-----ILKKQVEDRLKFTDTSAPMK 422  
QY 59 -----DGEWISGFEKKDAGVYNLSKDTPIKVPFKIEBKE---EENKPTF 103  
DB 423 NSDAIKAAALANGODIAGAEQKDVWV---SDEQVKEKEKEKEKEKKDKKEKK 479  
QY 104 DYSKKKQDPQVNHSQLNESHREKDLQREHSQKSDSTQ 142  
DB 480 DKREKDKKEKKDKK--DKRKSDDGEETPKKKKKKXKD 516  
XX  
RESULT 36  
AAY03189  
ID AAY03189 standard; Protein; 525 AA.  
XX  
AC AAY03189;  
XX  
DT 16-JUN-1999 (first entry)  
XX  
DE S. aureus trigger factor protein sequence.  
XX  
KM Trigger factor; tig; diagnosis; immune response; bacterial infection;  
KM S. aureus infection; otitis media; toxic shock syndrome; conjunctivitis;  
KM wound infection; impetigo; septic arthritis; therapy.  
XX  
OS Staphylococcus aureus.  
XX  
OS  
XX  
PN EP905233-A2.  
XX  
PD 31-MAR-1999.  
XX  
PF 21-AUG-1998; 98EP-0306697.  
XX  
PR 29-DEC-1997; 97US-0099339.  
PR 04-SEP-1997; 97US-0057511.  
XX  
XX (SMK ) SMITHKLINE BEECHAM CORP.  
XX  
PA (SMK ) SMITHKLINE BEECHAM PLC.  
XX  
PI Burnham MKR, Fosberry A, Hodgson JE, Jaworski DD;





```
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX (FARB ) BAYER AG.
XX
XX Tietjen K, Weidler M;
XX
XX MPI; 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5; SEQ ID NO 2552; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (AB90790-AB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 296 AA;
SQ
Query Match 10.5%; Score 89; DB 23; Length 296;
Best Local Similarity 23.1%; Pred. No. 2;
Matches 43; Conservative 33; Mismatches 70; Indels 40; Gaps 7;
QY 3 VKFPIINKGTEVSELPKRVVTYIIONK-----EMSGTIVSEEDTLP---- 46
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 26 ISQVALGESKNNTTE--PIQLYTVGSDKLLIGTSHKEKPPQSTETVIERNFALSHTWK 83
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 47 ----VYKGELEKGYQPDGMEISGFEKGKAGYVINISKQFIKPVFKTIEKKEENKPT 102
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 84 NSGVFSGYKPELIDQLNAAGFKGKGLLY-----PPAARKSAKQVNFQLPN 133
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 103 FDVSKKKDN---POVNHSQLNESHRKEDLOREHSOKSDSTQDVATVLDKNNISSKST 158
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 134 EDVKAKQDDADSDSESDSDSDSENSGDEEEKVTAESDSEEDSD--DEEDDSSEER 191
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 TNNPNK 164
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 192 T--PKK 195
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 39
ABB64828 standard; Protein; 281 AA.
XX
XX ABB64828;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 21276.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
```

```
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PMD, Myers EW;
XX
XX MPI; 2001-656860/75.
XX
XX N-PSDB; ABL08931.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 21276; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
XX sequences (AB101840-AB16175) and the encoded proteins
XX (AB16175-AB16175).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pot_sequences.
XX
XX Sequence 281 AA;
SQ
Query Match 10.4%; Score 88.5; DB 22; Length 281;
Best Local Similarity 22.1%; Pred. No. 2.1;
Matches 31; Conservative 34; Mismatches 52; Indels 23; Gaps 5;
QY 29 NKGEMSTIVSEEDFTLPYKGELEKGYQPDGMEISGFEKGKAGYVI-----NLS 79
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 137 NFEETFMNTTQODLKNLY-SLDR-----DNEGATSKELGVIRALGRPNESIA 188
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 80 KDTFIRPVFKTIEE-KKEENKPTEDVSKKKNPOVNHSQLNE-----SHRKEDLOREER 133
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 189 KEFECVILIRKHNDTKKEBELADAPRVFPKKNNGYISTELAAVFMALGKLEDEBLEEM 248
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 134 SOKSDSTQDVATVLDKNNI 153
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 249 IREYDLDQDNHINFEESNM 268
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 40
AAG85008 standard; Protein; 1141 AA.
XX
XX AAG85008;
XX
XX 11-SEP-2001 (first entry)
XX
XX Shrimp white spot Bacilliform virus (WSBV) protein 99.
XX
XX Shrimp white spot Bacilliform virus; MSB; diagnosis; viral infection;
XX antiviral agent; gene expression; antisense construct;
XX transgenic viral resistant shrimp.
XX
XX White spot syndrome virus.
XX
XX WO200138351-A2.
XX
XX 31-MAY-2001.
XX
XX 08-NOV-2000; 2000WO-US28888.
XX
XX 24-NOV-1999; 99CN-0124717.
XX
XX (PENY-) PE CORP NY.
XX
XX (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
XX
XX (SINO-) SINOGENOMAX CO LTD.
XX
```

XX Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;  
PI  
XX WPI; 2001-355877/37.  
DR N-PSDB; AAH62788.  
XX

PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus  
PT (WSBV), useful for producing viral polypeptides that can be used to  
PT screen for agents that are useful for treating WSBV infection -

PS Claim 1; Figure 3; 626pp; English.  
XX

CC The invention provides the primary nucleotide sequence of the WSBV genome  
CC (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and  
CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences  
CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid  
CC molecules and proteins of the invention are useful for diagnosis and  
CC monitoring viral infection, in screens for antiviral agents and for  
CC monitoring viral gene expression or activity during a treatment regimen.  
CC The nucleic acid molecules are also useful as antisense constructs to  
CC control viral gene expression in infected cells and tissues and to create  
CC transgenic viral resistant shrimp.

XX  
SQ Sequence 1141 AA;

Query Match 10.4%; Score 88.5; DB 22; Length 1141;  
Best Local Similarity 23.3%; Pred. No. 15;  
Matches 30; Conservative 31; Mismatches 43; Indels 25; Gaps 4;

QY 35 STIVSEEDPILPVYKGELEKGYQFDGMEISGPEG--KRDAGVYINLSKDTFIKPVFKIE 92  
DB 621 SNVEEBEE-----EEQMEEBEEEBEYEREBGSDKEDDGA-----PAQEME 662  
QY 93 EKKEEENKPTFVSKKQNPQVNHSQLNESHKEDLQREHSQKSDSTQDTATVLDKNN 152  
DB 663 EKKEEQQ-----QQPEEESNGNENQEEQQQQQPEREENKMDSDSDSSSSSSSS 717  
QY 153 ISSKSTNN 161  
DB 718 SSSSSSSSS 726

Search completed: February 10, 2004, 10:53:56  
Job time : 38.4525 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 7.62413 Seconds  
(without alignments)  
1011.574 Million cell updates/sec

Title: US-10-067-385-8\_Copy\_610\_773

Perfect score: 848  
Sequence: 1 TTVEFLINKDTGSEVSELPK.....ATVLDKNNISSKSTNNPNK 164

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97.5	11.5	1345	1 YH00 YEAST	P38800 saccharomyc
2	97.5	11.5	2468	1 MABP HUMAN	P46821 homo sapien
3	95	11.2	348	1 CYL2 HUMAN	O14093 homo sapien
4	95	11.2	1398	1 TOR2 PLARK	P41001 plasmodium
5	93.5	11.0	649	1 HS70 PYRSA	P37899 pyrenomonas
6	93	11.0	2464	1 MABP MOUSE	P14873 mus musculi
7	92.5	10.9	1888	1 YDR2 SCHPO	O14207 schizosach
8	92	10.8	1202	1 DPMO ASCIM	P22374 ascobolus i
9	92	10.8	1875	1 MUP1 YEAST	O02455 saccharomyc
10	91	10.7	893	1 GYRA STAEF	P54112 staphylococ
11	90.5	10.7	443	1 DACA BACSU	P08750 bacillus su
12	90.5	10.7	1332	1 SPY7 YEAST	P35177 saccharomyc
13	90	10.6	688	1 LIP STAEF	O02510 staphylococ
14	89.5	10.6	433	1 TIG STAEF	O09916 staphylococ
15	87.5	10.3	655	1 SKM1 YEAST	O12469 saccharomyc
16	87.5	10.3	1702	1 IGA2 HAEIN	P43584 haemophilus
17	87	10.3	778	1 YFR8 YEAST	P43610 haemophilus
18	87	10.3	6632	1 UN89 CAEEL	O011761 caenorhabdi
19	86.5	10.2	406	1 NSB1 MOUSE	O91135 mus musculi
20	86.5	10.2	1006	1 RAT1 YEAST	O02792 saccharomyc
21	86	10.1	443	1 GLNA YEAST	O08467 pyrococcus
22	86	10.1	914	1 PBPA BACSU	P39793 bacillus su
23	86	10.1	5596	1 MDN1 HUMAN	O9nnu22 homo sapien
24	85.5	10.1	752	1 DRS1 YEAST	P32892 saccharomyc
25	85.5	10.1	1007	1 RGAL YEAST	P39083 saccharomyc
26	85.5	10.1	2459	1 MABP RAT	P15055 rattus norv
27	85	10.0	719	1 YBLE SCHPO	O10342 schizosach
28	85	10.0	720	1 IF2 STAEF	O08624 staphylococ
29	84.5	10.0	657	1 Y040 MYCPN	P44963 mycoplasma
30	84.5	10.0	1694	1 IGA0 HAEIN	P44963 haemophilus
31	84	9.9	510	1 HG2V HALRO	O02508 haemophilus
32	84	9.9	1251	1 RBP2 PLAVB	O00799 plasmodium
33	83.5	9.8	360	1 FENR_PBA	P10933 pium sativ

34	83.5	9.8	479	1 U2R1 HUMAN	Q15695 homo sapien
35	83.5	9.8	678	1 GARP PLAVF	P13816 plasmodium
36	83.5	9.8	731	1 HS9A HUMAN	P07900 homo sapien
37	83.5	9.8	943	1 ARS2 DROME	O9y9k7 drosophila
38	83	9.8	258	1 IF31 HUMAN	O75822 homo sapien
39	83	9.8	270	1 TONB HAEIN	P42872 haemophilus
40	83	9.8	715	1 ADSV HUMAN	O096u3 homo sapien
41	83	9.8	847	1 DNLI CAEEL	O27474 caenorhabdi
42	83	9.8	3418	1 BRC2 HUMAN	P51587 homo sapien
43	82.5	9.7	439	1 GLNA PYRAB	O9uy99 pyrococcus
44	82.5	9.7	472	1 6PGD LACIA	O9cnu6 lactococcus
45	82.5	9.7	1553	1 TP2A CHICK	O04230 gallus gall

## ALIGNMENTS

RESULT 1	ID	YH00 YEAST	STANDARD	PRT	1345 AA.
AC	P38800	YH00 YEAST	STANDARD	PRT	1345 AA.
DT	01-FEB-1995	(Rel. 31, Created)			
DT	01-FEB-1995	(Rel. 31, Last sequence update)			
DT	01-NOV-1995	(Rel. 32, Last annotation update)			
DE	Hypothetical 149.7 kDa protein in IRR1-KSPL intergenic region.				
GN	YHR080C.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=S288C / AB972;				
RX	MEDLINE=94378003; PubMed=8091229;				
RA	Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,				
RA	Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kixeten J.,				
RA	Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,				
RA	Latreille P., Louis E.J., Macri C., Martis E., Menezes S., Mouser L.,				
RA	Nhan M., Rifkin L., Riles L., St Peter H., Trevaaks E., Vaughan K.,				
RA	Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,				
RA	Vaudin M.,				
RT	"Complete nucleotide sequence of Saccharomyces cerevisiae chromosome				
RT	VIII."				
RL	Science 265:2077-2082(1994).				
CC	-I- SIMILARITY: TO YEAST YFL042C.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	-----				
DR	EMBL; U10556; AAB68895.1; -				
DR	PIR; S46817; S46817.				
DR	SGD; S0001122; YHR080C.				
DR	InterPro; IPR004182; GRAM_dom.				
DR	Pfam; PF02893; GRAM_1.				
DR	SMART; SM00568; GRAM_1.				
KW	Hypothetical protein; Transmembrane.				
FT	TRANSMEM 1198 1218 POTENTIAL.				
FT	SEQUENCE 1345 AA; 149679 MW; 2FDAB94A66564C2 CRC64;				
SQ					
Query Match	11.5%; Score 97.5; DB 1; Length 1345;				
Best Local Similarity	27.0%; Pred. No. 11;				
Matches	33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;				
QY	49 KGELKGYQFDGWEISGEFGK-DAGYVINSKDTFKVFKIIEKKEENKPTFDVSK 107				
DB	1109 KGALIEK-----SVEGQKVSVDVWLSELRLDI-----SPAKSKKPKVKVM 1149				
QY	108 KKDNPQVNHSQLNSHREKEDIQREHSGKSDSTVDATATVD--KNNISSKSTN---NP 162				

Db 1150 SHKHPRFSKVE-----QKSSRSKSDNKKILTHLDVQNNFSEIEMNTLSP 1201

QY 163 NK 164

Db 1202 QK 1203

RESULT 2

MAPB\_HUMAN STANDARD; PRT; 2468 AA.

AC P46821;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Microtubule-associated protein 1B (MAP 1B) [contains: MAP1 light chain LC1].

DE MAP1B.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Fetal brain;

RX MEDLINE=95104835; PubMed=7806212;

RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;

RT "Cloning of human microtubule-associated protein 1B and the identification of a related gene on chromosome 15.";

RL Genomics 22:273-280(1994).

CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN. PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN STABILIZING MICROTUBULES.

CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE WITH MAP1B AND MAP1B PROTEINS.

CC -1- DOMAIN: Has a highly basic region with many copies of the sequence KKEE and KKEI/V, repeated but not at fixed intervals, which is responsible for the binding of MAP1B to microtubules.

CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated from MAP1B by proteolytic processing. It is free to associate with both MAP1A and MAP1B. It interacts with the amino-terminal region of MAP1B (by similarity).

CC -1- SIMILARITY: TO MAP1A.

CC -1- SIMILARITY: TO MAP1A.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; L06237; AAA18904.1; --

DR Genew; HGNC:6836; MAP1B.

DR MIM; 157129; --

DR GO; GO:0005875; C:microtubule associated complex; TAS.

DR InterPro; IPR000102; MAP1B neurxin.

DR Pfam; PF000414; MAP1B neurxin; 10.

DR PROSITE; PS00230; MAP1B NEURAXIN; 6.

KW Microtubules; Repeat; Phosphorylation.

FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.

FT REPEAT 1878 1894 MAP1B 1.

FT REPEAT 1895 1911 MAP1B 2.

FT REPEAT 1912 1928 MAP1B 3.

FT REPEAT 1929 1945 MAP1B 4.

FT REPEAT 1946 1962 MAP1B 5.

FT REPEAT 1963 1979 MAP1B 6.

FT REPEAT 1997 2013 MAP1B 7.

FT REPEAT 2014 2030 MAP1B 8.

FT REPEAT 2031 2047 MAP1B 9.

FT REPEAT 2048 2064 MAP1B 10.

FT DOMAIN 589 790 IYS-RICH (HIGHLY BASIC, CONTAINS MANY KKEE AND KKEI/V REPEATS).

SQ SEQUENCE 2468 AA; 270618 MW; 540839CDDP09D461 CRC64;

Query Match 11.5%; Score 97.5; DB 1; Length 2468;

Best Local Similarity 24.8%; Pred. No. 21;

Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 4 KEPIANKDTGEVSELPKPHRYVTYTIQNGKMSSTIYSEEDFILPYVKGELKGYFDGWEI 63

Db 584 EKVMVKKKPKVTEKTPSVTEKEVPSKEPS-----PV-KAEVA-----EK 623

QY 64 SEFEGKQAGYVINTSKDTFIKPVFKTEKEEKEEKPFPVSKKDNQVNHSQLNESH 123

Db 624 QATVKKPRAAKETVKKETKVP-----EDKKEEKEKPKVEVAKEDKTPI---KKEKP 675

QY 124 RKEDLQRE-----EHSQKSDSTKDV 143

Db 676 KKEVYKKEVKEIKKEEKEPKKEV 700

RESULT 3

CYL2\_HUMAN STANDARD; PRT; 348 AA.

AC 014093;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cylindrin II (Multiple-band polypeptide II).

GN CYL2 OR CYL2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RX MEDLINE=95255491; PubMed=7737358;

RA Hess H., Held H., Zimbelmann R., Franke W.W.;

RT "The protein complexity of the cytoskeleton of bovine and human sperm heads: the identification and characterization of cylindrin II.";

RL Exp. Cell Res. 218:174-182(1995).

CC -1- FUNCTION: POSSIBLE ARCHITECTURAL ROLE DURING SPERMATOGENESIS. MAY BE INVOLVED IN SPERMATID DIFFERENTIATION.

CC -1- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE.

CC -1- TISSUE SPECIFICITY: Testis.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; 246788; CAA86752.1; --

DR PIR; I37271; I37271.

DR Genew; HGNC:2583; CYL2.

DR MIM; 604035; --

DR GO; GO:0005200; P:structural constituent of cytoskeleton; TAS.

KW Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.

FT DOMAIN 25 347 31 X 3 AA REPEATS OF K-K-X.

FT REPEAT 157 240 3 X APPROXIMATE TANDEM REPEATS.

FT REPEAT 157 184 1.

FT REPEAT 185 212 2.

FT REPEAT 213 240 3.

SQ SEQUENCE 348 AA; 39079 MW; D86766599C1809E7 CRC64;

Query Match 11.2%; Score 95; DB 1; Length 348;

Best Local Similarity 30.9%; Pred. No. 3.9;

Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

```

QY 49 KGELEKGYQPDGMEISGFEGKKDAGVYINLSKDTFIKPFVKKEEKENKPTF---DV 105
DB 205 ESEBGEK---GTEKDSKKKKDS---KKGKDSKIEIQAAYKADKKEDGKKANKGDE 256
QY 106 SK--KKDNPOVNSQLN-----ESHKEDLOREHSQSDSTDD---VTATVLDKNNI 153
DB 257 SKDAKKDAKEIKKKKKKKKSDSTDSKDVKKY---SKDATDADKAVAKDOTEKESA 313
QY 154 SSK 156
DB 314 DSK 316

RESULT 4
TOP2_PLAFK STANDARD; PRT; 1398 AA.
ID TOP2_PLAFK STANDARD; PRT; 1398 AA.
AC P41001;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5839;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=94316496; PubMed=8041616;
RA Cheesman S., McAleese S., Goman M., Johnson D., Horrocks P.,
RA Ridley R.G., Kilbey B.J.;
RT "The gene encoding topoisomerase II from Plasmodium falciparum.";
RL Nucleic Acids Res. 22:2547-2551(1994).
CC -!- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -!- SIMILARITY: Belongs to the type II topoisomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X79345; -; NOT_ANNOTATED_CDS.
DR HSBP; P08786; 1EGW.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR003957; CBPA_NFYB_topis.
DR InterPro; IPR001241; DNA_topoisomII.
DR InterPro; IPR002205; DNA_topoisomIV.
DR Pfam; PF00204; DNA_gyraseb.1.
DR Pfam; PF00521; DNA_topoisomIV.1.
DR Pfam; PF02518; HATPase_C.1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PRINTS; PR00418; TP12FAMILY.
DR PRODOM; PD000742; DNA_topoisomIV.1.
DR SMART; SM00387; HATPase_C.1.
DR SMART; SM00433; TOP2C.1.
DR SMART; SM00434; TOP4C.1.
DR PROSITE; PS00177; TOPOISOMERASE_II.1.
DR Isoelectr; PS00177; TOPOISOMERASE_II.1.
DR Isoelectr; PS00177; TOPOISOMERASE_II.1.
DR NP_BIND 149 149 ATP (POTENTIAL).
DR ACT_SITE 830 830 DNA CLEAVAGE (BY SIMILARITY).
FT DOMAIN 271 281 POLY-ASN.
FT DOMAIN 308 316 POLY-ASN.

```

```

FT DOMAIN 1089 1093 POLY-LYS.
SQ DOMAIN 1327 1234 POLY-LYS.
SQ SEQUENCE 1398 AA; 161029 MW; BAAD7BEE86FE5BE9 CRC64;
Query Match
Best Local Similarity 23.8%; Pred. No. 17;
Matches 36; Conservative 36; Mismatches 53; Indels 26; Gaps 6;
QY 29 NGKEMSTIVSEEDPIL--PYKGELEKGYQPDGMEISGFEGKKDAGVYINLSKDTFIKPF 86
DB 1133 NEETIAGITVKDYDLSMPILFSLTEK--VEDLLTQLKERELEILRNITVETMWLK 1189
QY 87 VFKTIEE-----KKEEENKPTPDVSKKDNPOVNSQLNESHKEDLOREHSQ 135
DB 1190 DIEKVEAIEFQRVVELSNRESNK--FKVARKQ-----GRSSMKKKKKKLSDEESE 1242
QY 136 ---KSDSTKDYATVLDKNNISKSTTNP 163
DB 1243 GGDTSDSSEFLVNTLNKNTKTKTTSSNN 1273

RESULT 5
HS70_PYRSA STANDARD; PRT; 649 AA.
ID HS70_PYRSA STANDARD; PRT; 649 AA.
AC P37899;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heat shock 70 kDa protein.
GN HSP70.
OS Pyrenomonas salina.
OC Nucleomorph.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Pyrenomonas.
OX NCBI_TaxID=3034;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=94268506; PubMed=8208251;
RA Holman C.J.B., Remington S.A., Haebler M.M., Martin W.F., Mueller S.B.,
RA Couch J., McFadden G.I., Iglot G.L., Malar U.-G.;
RT "The smallest known eukaryotic genomes encode a protein gene: towards
RT an understanding of nucleomorph functions.";
RL Mol. Gen. Genet. 243:600-604(1994).
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X72621; CAA51197.1; -.
DR PIR: S42488; S42488.
DR HSBP; P08109; 1CKR.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70.1.
DR PRINTS; PR00301; HEATSHOCK70.
DR PRODOM; PD000089; Hsp70.1.
DR PROSITE; PS00297; HSP70_1.1.
DR PROSITE; PS00329; HSP70_2.1.
DR PROSITE; PS01036; HSP70_3.1.
DR ATP-binding; Heat shock; Nucleomorph.
SQ SEQUENCE 649 AA; 72079 MW; B627B08FF90C9164 CRC64;
Query Match
Best Local Similarity 24.5%; Pred. No. 9.7;
Matches 38; Conservative 26; Mismatches 50; Indels 41; Gaps 7;
QY 7 ILNKDTGVSEELKPRVTVTIQNG---KEMSTIVSEEDPILPYKGELEKGYQPDGMEI 63
DB 491 ILNVSASDKSTGSKSKNTITTDKGRSLKEELERWVEAE---KKTEDEK----- 537

```

QY 64 SGFEKGDQ-----GVINLS---KDTFLKPVFKIEKKEENKPTFVSKKKNPQVNA 115  
 DB 538 --LDRKLEKAKNSJENANVIRNRD-----EKAKIKQEDDKSIEKAYKE----- 582  
 QY 116 HSQLNESHRRKDLQREHSHQSDSTYDVTATVLDK 150  
 DB 583 --VLEFETNDELEKEEYKEKEKELNFPANPLISK 615

RESULT 6  
 MAPB\_MOUSE STANDARD; PRT; 2464 AA.  
 ID MAPB\_MOUSE  
 AC P14873;  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Microtubule-associated protein 1B (MAP1B) (MAP1.2) (MAP1(X))  
 DE [Contains: MAP1 light chain LC1].  
 GN MAP1B OR MTAP1B OR MTAP5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RP [1]  
 RP SEQUENCE FROM N.A., AND DOMAIN.  
 RX STRAIN=Swiss Webster; TISSUE=Brain;  
 RX MEDLINE=90094539; PubMed=2480963;  
 RA Noble M., Lewis S.A., Cowan N.J.;  
 RT "The microtubule binding domain of microtubule-associated protein  
 MAP1B contains a repeated sequence motif unrelated to that of MAP2  
 and tau.";  
 RL J. Cell Biol. 109:3367-3376(1989).  
 CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.  
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS  
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
 CC STABILIZING MICROTUBULES.  
 CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
 CC WITH MAP1B AND MAP1B PROTEINS.  
 CC -1- DOMAIN: Has a highly basic region with many copies of the sequence  
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is  
 CC responsible for the binding of MAP1B to microtubules.  
 CC -1- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED  
 CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH  
 CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION  
 CC OF MAP1B.  
 CC -1- SIMILARITY: TO MAP1A.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X51396; CAA35761.1; --  
 DR PIR; S07549; QRMSP1.  
 DR MGD; MG1:1306778; Mtap1b.  
 DR GO; GO:0016358; P:dendrite morphogenesis; IMP.  
 DR GO; GO:0001578; P:microtubule bundling; IMP.  
 DR InterPro; IPR000102; MAP1B\_neuraxin.  
 DR Pfam; PF00414; MAP1B\_neuraxin; 10.  
 DR PROSITE; PS00230; MAP1B\_NEURAXIN; 7.  
 DR Microtubules; Repeat; Phosphorylation.  
 FT CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.  
 FT REPEAT 1874 1890 MAP1B 1.  
 FT REPEAT 1891 1907 MAP1B 2.  
 FT REPEAT 1908 1924 MAP1B 3.  
 FT REPEAT 1925 1941 MAP1B 4.  
 FT REPEAT 1942 1958 MAP1B 5.  
 FT REPEAT 1959 1975 MAP1B 6.

FT REPEAT 1993 2009 MAP1B 7.  
 FT REPEAT 2010 2026 MAP1B 8.  
 FT REPEAT 2027 2043 MAP1B 9.  
 FT REPEAT 2044 2060 MAP1B 10.  
 FT DOMAIN 589 787 LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
 KKEE AND KKEI/V REPEATS).  
 SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFBDA87 CRC64;

Query Match 11.0%; Score 93; DB 1; Length 2464;  
 Best Local Similarity 27.8%; Pred No. 45;  
 Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;

QY 4 KEFLINKDTGSEVSEKPHRVTVTTONGKMSSTVSEDFILPVYKGELEKGYQFDGWEI 63  
 DB 584 EKVLVKKDKPVTKESKP---SVT---EKVVS---KEEG--SPV-KAEVA-----EK 623  
 QY 64 SGFEKGDQAGYINISKDTFLKPVFK-KKEEKEENKPTFVSKKKNPQVNHSLNS 122  
 DB 624 QATESKP-----KTKDKVVKKEIKTKLEEKKEE--KPKKVVKKEDTTPV---KDKK 672

QY 123 HRKEDLQRE-----EHSQKSDSTYDVTATVLDKNNISKSTNNPNK 164  
 DB 673 PKREYKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKEIKKE 728

RESULT 7  
 YD12\_SCHPO STANDARD; PRT; 1888 AA.  
 ID YD12\_SCHPO  
 AC 014207;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein C6B12.02c in chromosome I.  
 GN SPACB12.02C.  
 OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RX Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Frazer A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howard S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmons M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkhardt G., Aert R., Robben J., Grymoprez B.,  
 RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Meller-Auer S.,  
 RA Gabel K., Fuchs M., Fritze C., Holzer E., Moesl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lemnach H., Keimhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambut S., Purrelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux R., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Patahshin J.,  
 RA Shpakovski G.V., Useery D., Barrell B.G., Nurse P.;  
 RT The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (potential).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



```

or send an email to license@lsb-sib.ch.
-----
DR EMBL, X15982, CAA34106.1, -.
DR PIR, S05362, S05362.
DR InterPro, IPR006172, DNA_pol_B.
DR InterPro, IPR004868, DNA_pol_B_2.
DR Pfam, PF031175, DNA_pol_B_2, 1.
DR SMART, SMO0486, POLBc_1_1.
DR PROSITE, PS00116, DNA_POLYMERASE_B, 1.
KW transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Plasmid; Mitochondrion.
SQ SEQUENCE 1202 AA; 138279 MW; 51D41FCBDBF2CDE CRC64;

Query Match 10.8%; Score 92; DB 1; Length 1202;
Best Local Similarity 22.5%; Pred. No. 24;
Matches 42; Conservative 37; Mismatches 76; Indels 32; Gaps 9;

Qy 6 FILNK----PTGESELKPHRYVTYITONGKEMSGTI---VSEDF--ILPVYKGELEK-- 54
Db 311 FVNAAKIKFPFGNVRISIGFGNVT-TLTDKETLITLAIPLEREDIHVMASDDEIDDESK 369
Qy 55 ---GYOPDGEWEISGFEGKCDAGYVIMLSKDTFIKPVFKIE-----EKKEENK-PTF 103
Db 370 FPKGSLSEFDEKPLKTIEGTGKAYNTFPFKDIVKDIINKKINFGNLDLPKTMDSKWNPL 429
Qy 104 DVSKKKNPPVQNSQLMESHKEDLPQREHSQKSDSTKDVATVLDKNT-----SSK 156
Db 430 KLNDKTSIGELRMTRIKKNNOVDYI--IGHMINDGNVITFNRAVDNSIIRFTVTDSM 487
Qy 157 STTNPN 163
Db 488 GNTNDPN 494

RESULT 9
MLP1_YEAST STANDARD; PRT; 1875 AA.
ID MLP1_YEAST
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLP1.
GN MLP1 OR YKR035W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota, Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=S288C;
RC MEDLINE=93247549; PubMed=8483450;
RX Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369(1993).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=94205265; PubMed=8154186;
RX Bou G., Bateman P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballester J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
RT new open reading frames.";
RL Yeast 9:1349-1354(1993).
CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MLP1".
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```



RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaeser P., Goffeau A., Golightly E.U., Grandi G.,  
 RA Guisepi G., Guy B.U., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karmata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarovic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetle D., Rapoport G., Prescott A.M.,  
 RA Precourt E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi Y., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takuchi M., Tamakoshi A., Tanaka T., Terpeira P., Tognoni A.,  
 RA Toasato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambat R., Wedler E., Wedler H., Weitznegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasuno K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.,  
 RA "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RT Nature 390:249-256(1997).  
 RL [3]  
 RN SEQUENCE OF 32-102.  
 RP MEDLINE=80182289; PubMed=6768745;  
 RX Waxman D.J., Strominger J.L.;  
 RA "Sequence of active site peptides from the penicillin-sensitive D-  
 RT alanine carboxypeptidase of *Bacillus subtilis*. Mechanism of  
 RT penicillin action and sequence homology to beta-lactamases.";  
 RL J. Biol. Chem. 255:3964-3976(1980).  
 RN [4]  
 RN SEQUENCE OF 103-443 FROM N.A.  
 RP MEDLINE=86250602; PubMed=3087956;  
 RX Todd J.A., Roberts A.N., Johnstone K., Pigott P.J., Winter G.,  
 RA Ellier D.J.;  
 RT "Reduced heat resistance of mutant spores after cloning and  
 RT mutagenesis of the *Bacillus subtilis* gene encoding penicillin-binding  
 RT protein 5.";  
 RL J. Bacteriol. 167:257-264(1986).  
 RN [5]  
 RP SEQUENCE OF 414-443.  
 RX MEDLINE=81117303; PubMed=6780559;  
 RA Waxman D.J., Strominger J.L.;  
 RT "Primary structure of the COOH-terminal membranous segment of a  
 RT penicillin-sensitive enzyme purified from two *Bacilli*.";  
 RL J. Biol. Chem. 256:2067-2077(1981).  
 RN [1- FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE  
 CC CELL WALL PRECURSORS.  
 CC [1- CATALYTIC ACTIVITY: D-alanyl-D-alanine + H(2)O = 2 D-alanine.  
 CC [1- PATHWAY: Peptidoglycan synthesis; final stages.  
 CC [1- SUBCELLULAR LOCATION: Membrane-associated.  
 CC [1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11, ALSO KNOWN AS THE  
 CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 1 FAMILY.  
 CC [1-  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC [1-  
 CC EMBL, D26185; BAA05246.1; -  
 CC EMBL, Z99104; CAB11786.1; -  
 CC EMBL, M13766; AAA2375.1; -  
 CC PIR, S66040; S66040.  
 CC MEROPS, S11.001; -  
 CC Subtilast; BG10074; dact.  
 CC InterPro; IPR001967; Ala/AlaCbpase1.  
 CC Pfam; PF00768; Peptidase\_S11; 1.

DR PRINTS; PR00725; DADACBPASE1.  
 KW Hydrolyase; Carboxypeptidase; Peptidoglycan synthesis; Cell wall;  
 KM Membrane; Signal; Complete proteome.  
 FT SIGNAL 1 31  
 FT CHAIN 32 443 D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.  
 FT ACT SITE 67 67 ACYLATED BY PENICILLIN.  
 FT CONFLICT 100 100 E -> Q (IN REF. 3).  
 FT CONFLICT 227 227 E -> Q (IN REF. 4).  
 SQ SEQUENCE 443 AA; 48636 MW; DA6CSB0307DVC117 CMC64;  
 Query Match 10.7%; Score 90.5; DB 1; Length 443;  
 Best Local Similarity 24.1%; Pred. No. 11;  
 Matches 32; Conservative 25; Mismatches 49; Indels 27; Gaps 4;  
 QY 6 FLTKDTEVEELRPHRTVTIIONKEMSTIYSEEDPILRVYKGELEKYGQFQDMEISG 65  
 DB 310 FMSMEIYVGGQVGHK-TISVDKKEKEVGIVTKKASLPVKGGE-EKNYKAK----- 361  
 QY 66 FEGKADAGVYINLSKDTPIKVPFKIEKKEENKPTDVSKKDNPQVNHSHLSHRK 125  
 DB 362 -----VTLNKDNTATAVKKGITKVGK-----LTAETGDEKQYGFINSDLAG 402  
 QY 126 EDLQREHSQKSD 138  
 DB 403 VDLVTKEVKEKAN 415  
 RESULT 12  
 SPT7 YEAST STANDARD; PRT; 1332 AA.  
 ID SPT7 YEAST  
 AC P35177;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Transcriptional activator SPT7.  
 GN SPT7 OR YBR081C OR YBR0739.  
 OS *Saccharomyces cerevisiae* (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OK NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=95229044; PubMed=7713415;  
 RA Ganeheroff L.J., Dollard C., Tan P., Winston F.;  
 RT "The *Saccharomyces cerevisiae* SPT7 gene encodes a very acidic protein  
 RT important for transcription in vivo.";  
 RL Genetics 139:523-536(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=95076715; PubMed=7985423;  
 RA van der Aart Q.J.M., Barthe C., Dolignon F., Aigle M., Crouzet M.,  
 RA Steensma H.Y.;  
 RT "Sequence analysis of a 31 kb DNA fragment from the right arm of  
 RT *Saccharomyces cerevisiae* chromosome II.";  
 RL Yeast 10:959-964(1994).  
 RN [3]  
 RP SEQUENCE OF 1-835 FROM N.A.  
 RC STRAIN=S288C;  
 RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,  
 RA Visiers S.;  
 RT Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 463-523 FROM N.A.  
 RX MEDLINE=92285152; PubMed=1350857;  
 RA Haynes S.R., Dollard C., Winston F., Beck S., Trowsdale J.,  
 RA David I.B.;  
 RT "The bromodomain: a conserved sequence found in human, *Drosophila* and  
 RT yeast proteins.";  
 RL Nucleic Acids Res. 20:2603-2603(1992).  
 CC [1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY  
 CC OTHER GENES.

```
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 1 bromodomain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -1- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a
CC fatty acid anion.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS
CC NEGATIVELY REGULATED BY BACTERIOPHAGE LYSOGENIZATION (LIPASE
CC CONVERSION).
CC -1- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaico C.,  
 RA Setimizu K., Hiraoka H., Kuhara S., Goto S., Yabuzaki J.,  
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,  
 RT "Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*,"  
 RL Lancet 357:1225-1240(2001).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MM2;  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Ikawa N., Asano K., Naito T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiramatsu K.,  
 RT "Genome and virulence determinants of high virulence community-acquired MRSA,"  
 RL Lancet 359:1819-1827(2002).  
 CC -1- FUNCTION: Involved in protein export. Acts as a chaperone by maintaining the newly synthesized protein in an open conformation (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY. TIG SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AP003363; BAB57837.1; -  
 DR EMBL: AP003134; BAB42766.1; -  
 DR EMBL: AP004827; BAB95484.1; -  
 DR PIR: A89951; A89951.  
 DR HAMAP: MF\_00303; -; 1.  
 DR InterPro: IPR001179; FKBP\_PPIase.  
 DR InterPro: IPR005215; Trig\_fac.  
 DR Pfam: PF00254; FKBP; 1.  
 DR TIGRfam: TIGR00115; Lig; 1.  
 DR PROSITE: PS00453; FKBP\_PPIASE\_1; FALSE\_NEG.  
 DR PROSITE: PS00454; FKBP\_PPIASE\_2; FALSE\_NEG.  
 DR PROSITE: PS50059; FKBP\_PPIASE\_3; 1.  
 KM Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.  
 FT DOMAIN 163 248 PPIASE, FKBP-TYPE.  
 SQ SEQUENCE 433 AA; 48609 MW; 8865D9AF6A1BC1E7 CRC64;  
 Query Match 10.6%; Score 89.5; DB 1; Length 433;  
 Best Local Similarity 22.1%; Pred. No. 12; Indels 83; Gaps 13;  
 Matches 50; Conservative 33; Mismatches 60;  
 QY 11 DTGEVSELKPRVTVT-IQNGKE--MSGTIVSEEDFLLPVYKG-ELEK----- 54  
 DB 81 DETDKPVAAQPEVSVTQIEKGDFFPAIVTVPEPVKGDYGLIEKQETLSDELQ 140  
 QY 55 -----GYOFGWEIISG--FESKQAGVYINISKQTFIKP 86  
 DB 141 AIDHSIGLAEVWVEDGVENGDTVINIDFSG-SVDSGEFEGQAGEVDLEIGSGSFI-P 198  
 QY 87 VFK-----KIEEKE-----EK-----KKKNPQNNHSQLAE- 121  
 DB 199 GFEOLBGMKQDEKQVAVTTPPEYHAELACKKATFTKKNVIEFKPEVPELTDEIANEL 258  
 QY 122 -----SHRKEDLQREHSHQSKSDTKDVTATVLDKNNISKSTTN 160  
 DB 259 DAEANTVDEYKENLKRLLAEQKATDAENV-----EKEKATITATQN 299  
 RESULT 15  
 SKM1\_YEAST STANDARD; PRT; 655 AA.  
 AC 012469; 006940;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Serine/threonine-protein kinase SKM1 (EC 2.7.1.1-) (Protein kinase  
 DE 75490 D).  
 GN SKM1 OR YOL113W OR HRA655.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OC NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / GRF88;  
 RX MEDLINE=97197189; PubMed=9044278;  
 RA Martin H., Mendoza A., Rodriguez-Pachon J.M., Molina M., Nombela C.,  
 RT "Characterization of SKM1, a Saccharomyces cerevisiae gene encoding a novel Ste20/PAK-like protein kinase,"  
 RL Mol. Microbiol. 23:431-444(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96076631; PubMed=7502582;  
 RA Vandenbol M., Durand P., Portetelle D., Hilger F.,  
 RT "Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV including the TYI-H3 retrotransposon, the sufl(+ ) frameshift RT suppressor gene for tRNA-Gly, the yeast transfer RNA-Thr-1a and a delta element,"  
 RL Yeast 11:1069-1075(1995).  
 CC -1- FUNCTION: MAY BE INVOLVED IN CELLULAR SIGNALING OR CYTOSKELETAL  
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC STE20 SUBFAMILY.  
 CC -1- SIMILARITY: Contains 1 CRIB domain.  
 CC -1- SIMILARITY: Contains 1 PH domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X69322; CAA49163.1; -  
 DR EMBL: Z48149; CAA88147.1; -  
 DR EMBL: Z74855; CAA99132.1; -  
 DR PIR: S51884; S51884.  
 DR SGD: S0005473; SKM1.  
 DR GO: GO:000074; P:regulation of cell cycle; IGI.  
 DR InterPro: IPR000095; PAKbox/Rho-binding.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR000719; Prot\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00786; PBD; 1.  
 DR Pfam: PF00169; PH; 1.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Prot\_kinase; 1.  
 DR SMART: SMO0285; PBD; 1.  
 DR SMART: SMO0233; PH; 1.  
 DR SMART: SMO0220; S\_TGK; 1.  
 DR PROSITE: PS50108; CRIB; 1.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 KM Transferase; Serine/threonine-protein kinase; ATP-binding.  
 FT DOMAIN 3 118  
 FT DOMAIN 123 136  
 FT DOMAIN 360 639  
 FT NP\_BIND 366 374 ATP (BY SIMILARITY);  
 FT BINDING 406 406 ATP (BY SIMILARITY).  
 FT ACT\_SITE 507 507 BY SIMILARITY.  
 FT CONFLICT 303 303 S -> A (IN REF. 1).  
 FT CONFLICT 306 306 R -> K (IN REF. 1).  
 FT CONFLICT 320 320 H -> Y (IN REF. 1).  
 FT CONFLICT 320 320



DR SCD: S0001934; YF0038W.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR000330; SNE2\_N.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00176; SNE2\_N; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC; 1.  
 DR Hypothetical protein; Nuclear protein; DNA-binding; Helicase;  
 KM ATP-binding.  
 FT NP BIND 247 254 ATP (POTENTIAL).  
 FT SITE 352 355 DECH BOX.  
 SQ SEQUENCE 778 AA; 88730 MW; 356C0857B5EAD84 CRC64;  
 Query Match 10.3%; Score 87; DB 1; Length 778;  
 Best Local Similarity 24.2%; Pred. No. 35;  
 Matches 44; Conservative 26; Mismatches 70; Indels 42; Gaps 6;  
 QY 4 KEFLINKDTGSELSKPRHVTYITNGKEMST-----IVSEDFILPYKGELEKG 55  
 DB 35 RKKEVNDLTLADISDSDDLSDSEDKHGKNDTAPITWLODDVHSDDI----- 82  
 QY 56 YQPDGMEISGEFGKADAGVNIISKDTPIKPVFKIEKEEENKPTFVSKKDPQVN 115  
 DB 83 -QLDSEDSDTFNAVQ-AQVVDKLADT--KSFQKSLDELSDMDIKTVSLAKKINFEVR 138  
 QY 116 HSQINES-----HRKEDLQR-----EEHSQKSDTKDVTATVLDKNNISSKS 157  
 DB 139 OSQVYSIIADTLTLRSNEVANANTKONSNDSEHSSKRRKTKKSLTFDFKQKKMED 198  
 QY 158 TT 159  
 DB 199 TT 200  
 RESULT 18  
 UN89\_CABEL STANDARD; PRT; 6632 AA.  
 ID UN89\_CABEL  
 AC 001761; Q17362;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).  
 GN UNC-89 OR C09D1.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.  
 RC STRAIN-Bristol N2;  
 RX MEDLINE=96180278; PubMed=8603916;  
 RA Benjan G.M., Tinley T.L., Tang X., Borodovsky M.;  
 RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line  
 assembly, encodes a giant modular protein composed of Ig and signal  
 transduction domains.";  
 RT J. Cell Biol. 132:835-848 (1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RA Du Z., Le T.T., Wilson R.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP REVISIONS.  
 RA Waterston R.;  
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Structural component of the muscle M-line. Myofibril  
 latic assembly begins with positional cues laid down in the  
 basement membrane and muscle cell membrane. UNC-89 responds to  
 these signals, localizes, and then participates in assembling an  
 M-line.  
 CC -1- TISSUE SPECIFICITY: Localizes to the middle of A-bands.  
 CC -1- SIMILARITY: Contains 1 DBL-homology (DH) domain.

CC -1- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -1- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 1 PH domain.  
 CC -1- SIMILARITY: Contains 5 RCDSD domains.  
 CC -1- SIMILARITY: Contains 1 SH3 domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb.ch/announce/>  
 or send an email to [license@isb.ch](mailto:license@isb.ch)).  
 CC EMBL; U33058; AAB00542.1; -.  
 CC EMBL; AF003131; AAB54132.2; -.  
 CC PDB; 1PHO; 20-DEC-00.  
 DR WormPep; C09D1.1; CE30426.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR007110; Ig\_1like.  
 DR InterPro; IPR003598; Ig\_c2.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR01849; PH.  
 DR InterPro; IPR007850; RCDSD.  
 DR InterPro; IPR000219; RhGEF.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00041; Im3; 1.  
 DR Pfam; PF00047; Ig; 47.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF05177; RCDSD; 5.  
 DR Pfam; PF00621; RhGEF; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR SMART; SM00408; IGC2; 23.  
 DR SMART; SM00325; RhGEF; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS50010; DH\_2; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 49.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;  
 KW 3D-structure.  
 FT DOMAIN 63 127 SH3.  
 FT DOMAIN 152 330 DH.  
 FT DOMAIN 342 498 PH.  
 FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.  
 FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.  
 FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.  
 FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.  
 FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.  
 FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.  
 FT DOMAIN 1272 1315 THR-RICH.  
 FT DOMAIN 1375 1475 RCDSD 1.  
 FT DOMAIN 1479 1585 RCDSD 2.  
 FT DOMAIN 1597 1695 RCDSD 3.  
 FT DOMAIN 1700 1799 RCDSD 4.  
 FT DOMAIN 1800 1860 RCDSD 5.  
 FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.  
 FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.  
 FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.  
 FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.  
 FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.  
 FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.  
 FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.  
 FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.  
 FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.  
 FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.  
 FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.  
 FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.  
 FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.  
 FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.  
 FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.  
 FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.  
 FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.



Query Match	Best Local Similarity	Score	DB 1:	Length	6632:
Matches	37;	Conservative	33;	Mismatches	80;
				Indels	20;
				Gaps	7;
1	TTVKEFLINKDTGEVSELKPHRVATYIIONGKMSSTIVSEEDF--ILPYYKGELKGYQF	58			
Db	1127	TKAKLTQKPKPAPFDFLR--VSLITVKGSE--AVFSAAHFGILPITYENSVAGKRV	1181		
Qy	59	DGWE--ISGPEKGDAGVIVNLISKOTFIKPV---FKKIEBK--EENKPTPDVSKK	109		
Db	1182	DQEGARVYTRDSEYTDGASILITDIPATYYSEVNHLTISVAENLTGAEITGQGLTIEPK	1241		
Qy	110	DNPOVNHQNLNESHKEDLOREHSHQKSDSTYQDVATYALDKNNISKSTT	159		

1242 ESUYVEKQDLS\$-----EVQKEIAQOVKASEPATTIMETSLSTKTT 1287

RESULT 19

ID	NSBI_MOUSE	STANDARD	PRT	406 AA.
AC	09JL35	088632; 08VC71; 09CUM1;		
DT	28-FEB-2003	(Rel. 41, Created)		
DT	28-FEB-2003	(Rel. 41, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	Nucleosome binding protein 1 (Nucleosome binding protein 45) (NBP-45) (GARP45 protein).			
GN	NSBP1 OR GARP45.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OK	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE			
RP	SPECIFICITY, AND DEVELOPMENTAL STAGE.			
RX	SHIRAKAWA H., Landman D., Postnikov Y.V., Bustin M.;			
RA	"NBP-45, a novel nucleosomal binding protein with a tissue-specific			
RT	and developmentally regulated expression."			
RL	J. Biol. Chem. 275:6368-6374(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	Onoda G., Suzuki N., Saito H., Honda T., Sato H., Kuwano R.;			
RL	Submitted (Sep-1998) for the EMBL/Genbank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Colon;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschuler S.F., Ziegler B., Buetow K.H., Scheffer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Dietchenko L., Marzita K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo W.F., Caesavert T.L., Schetz T.E.,			
RA	Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Rata S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Wozley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,			
RA	Villalona D.K., Muzny D.M., Sodeggen E.J., Lu X., Gibbs R.A.,			
RA	Faley J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting M., Maden A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Rodriguez A.C., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Butterfield Y.S.N., Krzywnicki M.I., Skalska U., Smalls D.E.,			
RA	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RN	[4]			
RP	SEQUENCE OF 7-195 FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=hippocampus;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi T., Fukuda S.,			
RA	Aitawa K., Iizawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,			
RA	Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Caavaent T.,			
RA	Fleischmann W., Gaasterland T., Glassi C., King B., Koehne H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikido I., Pesole G., Quackenbush J.,			
RA	Schirral L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,			
RA	Blake Y., Botfelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Guatinchich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima T., Mazzarelli J., Nombaeerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toy-o-oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,			

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayashizaki Y.,  
RT "Functional annotation of a full-length mouse cDNA collection."  
RL Nature 409:685-690(2001).  
CC -1- FUNCTION: Binds specifically to nuclear core particles and acts as  
CC a transcriptional activator.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: Widely expressed with highest levels in  
CC submaxillary gland, thymus, kidney and liver and lowest levels in  
CC brain, lung, pancreas and eye.  
CC -1- DEVELOPMENTAL STAGE: Highest levels are found in 7-day-old  
CC embryos. Levels in the 7-day-old embryo are 4-fold higher than in  
CC the adult and almost 10-fold higher than in later embryonic  
CC stages.  
CC -1- SIMILARITY: BELONGS TO THE HMG-14/HMG-17 PROTEIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AF213454; AAF30179.1; -;  
CC EMBL: AB018374; BAA33783.2; -;  
CC EMBL: BC021626; AAB21626.1; -;  
CC EMBL: AK013748; BAB28982.1; -;  
CC GDI; MG1:1355295; Ndbp1.  
CC GO: GO:0005654; C:nucleoplasm; IDA.  
CC GO: GO:0003682; F:chromatin binding activity; IDA.  
CC GO: GO:0006356; P:regulation of transcription from Pol I prom. .; IDA.  
CC InterPro: IPR000079; HMG\_14\_17.  
CC Pfam: PF01101; HMG14\_17; 1.  
CC SMART: SM00527; HMG17; 1.  
CC PROSITE: PS00355; HMG14\_17; FALSE NEG.  
CC DR Transcription regulation; Activator; DNA-binding; Nuclear protein.  
CC DR CONFLICT 74 74 M -> V (IN REF. 2).  
CC FT CONFLICT 390 390 N -> H (IN REF. 1).  
CC SQ SEQUENCE 406 AA; 45344 MW; 5944305613EC9679 CRC64;  
CC -----  
CC Query Match 10.2%; Score 86.5; DB 1; Length 406;  
CC Best Local Similarity 28.0%; Pred. No. 19;  
CC Matches 44; Conservative 16; Mismatches 52; Indels 45; Gaps 8;  
CC -----  
CC QY 9 NKTGEVSELKPRVTITQNGK-----ENSTIVSEEDFILPYKGELEKGYQFDGW 61  
CC DB 141 HKDTGE-----EVEDGKIEBEGLEKPRGAKSBD---AEVSKDEBKGNKNG- 185  
CC QY 62 EISGEGKKDAGVYINLSKD-----TFIKPYPKKIEEK---KEENKPTDVSKK 108  
CC DB 186 -----EDGEDEDEKEEDKDEGDTGEKEVEKONKEAEEDGCKKEENK---EVGKE 237  
CC QY 109 KDNPOVNSQLNESHKEDLQRE--HSQKSDSTKDV 143  
CC DB 238 GQPEEDGKEDLHEVGEKEDLHEDEGKEGQPEEDGKEI 274  
CC -----  
CC RESULT 20  
CC RAT1 YEAST  
CC ID RAT1 YEAST STANDARD; PRT; 1006 AA.  
CC AC 002792;  
CC DT 01-OCT-1993 (Rel. 27, Created)  
CC DT 01-OCT-1993 (Rel. 27, Last sequence update)  
CC DT 15-SEP-2003 (Rel. 42, Last annotation update)  
CC DE Ribonucleic acid trafficking protein 1 (5'-3' exoribonuclease)  
CC (EC 3.1.11.-) (P116)  
CC GN RAT1 OR HKEL OR TRAP1 OR YOR048C.  
CC OS Saccharomyces cerevisiae (Baker's yeast).  
CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
CC CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
CC OX NCBI\_TaxId=4932;  
CC RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=92331925; PubMed=1628825;  
RA Ambler D.C.; Goldstein A.L.; Cole C.N.;  
RT "Isolation and characterization of RAT1: an essential gene of  
RT Saccharomyces cerevisiae required for the efficient nucleocytoplasmic  
RT trafficking of mRNA."  
RL Genes Dev. 6:1173-1189(1992).  
RN [2]  
RP SEQUENCE FROM N.A. AND EXORIBONUCLEASE ACTIVITY.  
RX MEDLINE=93109318; PubMed=8417335;  
RA Kenna M., Stevens A., McCammon M., Douglas M.G.;  
RT "An essential yeast gene with homology to the exonuclease-encoding  
RT XRN1/XEM1 gene also encodes a protein with exoribonuclease  
RT activity."  
RL Mol. Cell. Biol. 13:341-350(1993).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93268292; PubMed=8497260;  
RA Aldrich T.L., di Segni G., McConaughy B.L., Keen N.J., Whelen S.,  
RA Hall B.D.;  
RT "Structure of the yeast TRAP1 protein: dependence of transcription  
RT activation on the DNA context of the target gene."  
RL Mol. Cell. Biol. 13:3434-3444(1993).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Landt O., Hiesel R., Unsel M.;  
RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Bohm C., Bolotin-Fukuhara M., Daignan-Fornier B., Dang D.V.,  
RA Valens M.;  
RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.  
CC CC  
CC -1- FUNCTION: May function in the processing and/or trafficking of  
CC nuclear mRNA. May be involved in general transcription as well.  
CC Possesses 5'->3' exoribonuclease activity degrading poly(A) to  
CC mainly 5'-AMP.  
CC -1- COFACTOR: Requires magnesium.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: Belongs to the 5'-3' exonuclease family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: S61567; AAB26818.1; -;  
CC EMBL: M95626; AAA34860.1; -;  
CC EMBL: L06011; AAA16950.1; -;  
CC EMBL: Z11746; -; NOT ANNOTATED CDS.  
CC EMBL: Z74956; CAA9240.1; -;  
CC PIR: S20126; S20126.  
CC SGD: S0005574; RAT1.  
CC GO: GO:0005634; C:nucleus; IDA.  
CC GO: GO:0004534; F:5'-3' exoribonuclease activity; IDA.  
CC GO: GO:0006365; P:35S primary transcript processing; IMP.  
CC GO: GO:0006396; P:RNA processing; IMP.  
CC InterPro: IPR004859; Put\_53exo.  
CC Pfam: PF01159; XRN\_N\_1; 1.  
CC KM Nuclear protein; Hydroxylase; Nuclease; Exonuclease; Repeat.  
CC FT DOMAIN 525 528 POLY-GLU.  
CC FT DOMAIN 793 797 POLY-ASN.  
CC FT DOMAIN 955 999 CONTAINS 2 X SRVD, 2 X NNNY, AND 2 X  
CC MUTAGEN 683 683 YSGN REPEATS.  
CC Y->H. IN ALLELE TRAP1-1; ACTIVATES  
CC TRANSCRIPTION OF THE PROMOTER-DEFECTIVE  
CC YEAST SUP4 TRNA(TYR) ALLELE SUP4A53T61.  
CC FT FT  
CC SQ SEQUENCE 1006 AA; 115933 MW; 5DD5B0245F3E12A CRC64;  
CC -----  
CC Query Match 10.2%; Score 86.5; DB 1; Length 1006;  
CC Best Local Similarity 19.0%; Pred. No. 50;  
CC -----

Matches 32; Conservative 40; Mismatches 47; Indels 49; Gaps 8;

QY 27 IONKMSSTIYSEDFILPVY--KELEKGYFDCWEISGFGKADAGVNLSDTPI 84  
 Db 415 MSKGDHRHPLVATEQ--LQMYDTQGNLAKG---SMNLTTSD-----MYRLKKEML 460  
 QY 85 -----KRVPKRIEKEEKEEKKPFEDVSKKKDNFQVHNSQ-----LINESHRK 125  
 Db 461 ANEGNEBAIAKVKQSDKNNELMKDISKEEIDAVSKANKTNPNLAEWKQKITINKKHL 520

QY 126 EDLOREH-----SOKSDSTKDVATVLDK-----NNISKSSTT 159  
 Db 521 EKONEEBEIAKSKYKTKAESECDLDEIKDEIYADVNDRENSETT 568

RESULT 21  
 GLNA PYRKO STANDARD; PRT; 443 AA.  
 AC 008467;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
 GN GLNA.  
 OS Pyrococcus kodakarensis.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Thermococcus.  
 OX NCBI\_TaxID=69014;  
 RN RP SEQUENCE FROM N.A.  
 RC STRAIN=KOD1.  
 RX MEDLINE=97316461; PubMed=9172372;  
 RA Adul Rahman K.N.Z., Jongsareejit B., Fujiwara S., Imanaka T.;  
 RT "Characterization of recombinant glutamine synthetase from the  
 RT hyperthermophilic archaeon Pyrococcus sp. strain KOD1.";  
 RL Appl. Microbiol. 63:2472-2476(1997).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +  
 CC L-glutamine.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D86222; BAA20530.1; -.  
 DR HSSP; P06201; IUGR.  
 DR InterPro; IPR001691; GLN synth.  
 DR InterPro; IPR004809; GLNA.  
 DR InterPro; IPR001637; GLNA adenyltn.  
 DR Pfam; PF00120; gln-synt; 1.  
 DR Pfam; PF03951; gln-synt\_N; 1.  
 DR ProDom; PD001057; GLN\_synt\_C; 1.  
 DR TIGRFAMs; TIGR00653; GLNA\_1; 1.  
 DR PROSITE; PS00180; GLNA\_1; 1.  
 DR PROSITE; PS00181; GLNA\_ATP; 1.  
 KM ligase.  
 FT BINDING 362 362 AMP (UNDER CONDITIONS OF ABUNDANT  
 FT GLUTAMINE) (BY SIMILARITY).  
 SQ SEQUENCE 443 AA; 50259 MW; 9426DCCEFE18168 CRC64;

Query Match 10.1%; Score 86; DB 1; Length 443;  
 Best Local Similarity 17.8%; Pred. No. 22;  
 Matches 31; Conservative 23; Mismatches 46; Indels 72; Gaps 4;

QY 48 YKGELEKGYFDCWEISGFGKADAGVNLSDTPI----- 84  
 Db 42 YEAAVEGVGSDSSIFGFGIDSDILFRADSTYALIEIWEGIGRYGYTKGDEPYQA 101

QY 85 --KRVFKRIEKEEKEEEN-----KPTFVSKKCD----- 110  
 Db 102 DRGLIKVLERLERLEKKAHNGPEPEFYIFKKNQWELHLPDSGCVFDVLGDKAREIR 161  
 QY 111 -----NPQVNSQLNESHRKEDLOREHSQSD---STKDVATVLD 149  
 Db 162 REIALYMPYLGLKPEVLHHEVGAKQHEIDFRDEALRTADNIVSPKRVKAVAE 215

RESULT 22  
 BBPA BACSU STANDARD; PRT; 914 AA.  
 ID BBPA BACSU  
 AC P39793;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Penicillin-binding protein 1A/1B (PBP1) (includes: Penicillin-  
 DE insensitive transglycosylase (EC 2.4.2.-) (peptidoglycan Tgase));  
 DE Penicillin-sensitive transpeptidase (EC 3.4.-.-) (DD-transpeptidase)].  
 GN PONA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN RP SEQUENCE FROM N.A., AND SEQUENCE OF 499-515.  
 RC STRAIN=168;  
 RX MEDLINE=95113769; PubMed=7814321;  
 RA Popham D.L., Setlow P.;  
 RT "Cloning, nucleotide sequence, and mutagenesis of the Bacillus  
 RT subtilis pona operon, which codes for penicillin-binding protein  
 RT (PBP) 1 and a PBP-related factor.";  
 RL J. Bacteriol. 177:326-335(1995).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=168 / Marburg;  
 RX MEDLINE=96349105; PubMed=8760912;  
 RA Sorokin A.V., Azevedo V., Zumbstein E., Galleron N., Ehrlich S.D.,  
 RA Serror P.;  
 RT "Sequence analysis of the Bacillus subtilis chromosome region between  
 RT the *serA* and *kds1* cloned in a yeast artificial chromosome.";  
 RL Microbiology 142:2005-2016(1996).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunes F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Betteiro M.G., Bessieres P., Bolotin A., Borcherdt S.,  
 RA Bortiss R., Boursier L., Brans A., Braun M., Britnell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Eutlan K.D., Errington J., Fabret C., Ferrati E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallazzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Goldlighty E.J., Grandi G.,  
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Perowalik S., Prescott A.M.,  
 RA Presecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivoita C., Roche E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiuchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taccoci E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandendol M., Vannier F., Vasseroiti A.,  
 RA Viari A., Wambutt R., Wedler R., Wedler H., Welltsegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;

RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 subsp.":  
 RL Nature 390:249-256 (1997).  
 [4]  
 RP GROWTH REQUIREMENTS.  
 RC STRAIN=168;  
 RX MEDLINE=98389671; PubMed=9721295;  
 RA Murray T., Popham D.L., Setlow P.;  
 RT "Bacillus subtilis cells lacking penicillin-binding protein 1 require  
 increased levels of divalent cations for growth.";  
 RL J. Bacteriol. 180:4555-4563 (1998).  
 [5]  
 RN SUBCELLULAR LOCATION.  
 RC STRAIN=168;  
 RX MEDLINE=99255546; PubMed=10322023;  
 RA Pedersen L.B., Angert E.R., Setlow P.;  
 RT "Septal localization of penicillin-binding protein 1 in *Bacillus*  
 subsp.":  
 RL J. Bacteriol. 181:3201-3211 (1999).  
 CC -1- FUNCTION: CELL WALL FORMATION. SYNTHESIS OF CROSS-LINKED  
 PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A  
 PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN  
 (FORMATION OF LINEAR GLYCANS) AND A PENICILLIN-SENSITIVE  
 TRANSPEPTIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE  
 SUBUNITS) (BY SIMILARITY).  
 CC -1- PATHWAY: Peptidoglycan synthesis; final stages.  
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. PROBABLY FOUND ALL  
 OVER THE WHOLE CELL AT LOW CONCENTRATIONS. ALSO LOCALIZES TO THE  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS CONSTANT DURING GROWTH,  
 DECREASES DURING SPOULATION AND IS INDUCED APPROXIMATELY 15 MIN  
 INTO SPORE GERMINATION.  
 CC -1- PTM: THE PRODUCT EXPRESSED FROM THE TRANSLATION OF THE PONA GENE  
 APPEARS AS TWO BANDS ON A GEL (1A AND 1B), BUT THE SPECIFIC AMINO  
 ACID SEQUENCE OF EACH PROTEIN IS UNKNOWN.  
 CC -1- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -1- MISCELLANEOUS: CELLS LACKING THE PROTEIN REQUIRE INCREASED LEVELS  
 OF MG(2+) OR CA(2+) FOR GROWTH AND GERMINATION. APPROXIMATELY 50%  
 OF CELLS WITHOUT THE PROTEIN CONTAIN ABNORMAL FT2Z RINGS,  
 SUGGESTING IT IS INVOLVED IN SEPTUM SYNTHESIS; INCREASED LEVELS OF  
 MG(2+) OR CA(2+) ONLY PARTIALLY ELIMINATE THE SEPARATION DEFECTS.  
 CC -1- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE  
 TRANSGLYCOSYLASE FAMILY.  
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE  
 TRANSPEPTIDASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, U11883; AAA64947.1; -;  
 DR EMBL, L47838; AAB38459.1; -;  
 DR EMBL, Z99115; CAB14148.1; -;  
 DR PIR, I40529; I40529.  
 DR Subtilisin, BG10954; PONA.  
 DR InterPro, IPR003961; FN\_III.  
 DR InterPro, IPR001264; Glyco\_trans\_51.  
 DR InterPro, IPR001460; Transpeptidase.  
 DR Pfam, PF00041; fn3; 1.  
 DR Pfam, PF00912; Transglycosyl; 1.  
 DR Pfam, PF00905; Transpeptidase; 1.  
 DR ProDom, PD001895; Glyco\_trans\_51; 1.  
 DR SMART, SM00060; FN3; 1.  
 DR Peptidoglycan synthetase; Cell wall; Transferase; Glycosyltransferase;  
 KW Hydrolase; Multifunctional enzyme; Transmembrane; Signal-anchor;  
 KW Antibiotic resistance; Complete proteome.  
 FT DOMAIN 1 37 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 38 58 SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN)

FT FT (POTENTIAL).  
 FT DOMAIN 59 914 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 77 246 TRANSGLYCOSYLASE.  
 FT DOMAIN 329 662 TRANSPEPTIDASE.  
 FT DOMAIN 706 784 FIBRONECTIN TYPE-III.  
 FT ACT SITE 390 390 ACYLATED BY PENICILLIN (BY SIMILARITY).  
 SQ SEQUENCE 914 AA; 99562 MW; 6978E33DFE2423E6 CRC64;  
 Query Match 10.1%; Score 86; DB 1; Length 914;  
 Best Local Similarity 23.0%; Pred. No. 49; Indels 62; Gaps 9;  
 Matches 45; Conservative 33; Mismatches 56; Indels 62; Gaps 9;  
 QY 3 VEFPLNDTGEVSEELKPHRTVTITQNGKMSSTIVSEDFILPYKSELEKGYQPDGE 62  
 DB 737 VKQSV---DGGSYSB-----IQNS-----AEAVISGVQPSV---YKFEVTA 774  
 QY 63 ISGFEKKDA--GYVINSKDTFKVPEKLEKKEEENK----- 100  
 DB 775 VSD-DGKSTASTSYEVPFAEDDEDKDDQQTDEKQDDTQDTQDSQKDDQDQD 833  
 QY 101 ----PTFVSKKKNDPQVNHSQLN-----ESHRKEDLQREHSQKSDSTKDYTA 145  
 DB 834 QTDSTNDQDKQDDTNTNPSDNNQDOSNDNDNSNNQDTSQDSNGKNDSTG--SD 891  
 QY 146 TVLDGNISSKSTNN 161  
 DB 892 TNKNTDTSNKQTINS 907  
 RESULT 23  
 MDN1\_HUMAN STANDARD; PRT; 5596 AA.  
 ID MDN1\_HUMAN  
 AC Q9NUZ2; O15019;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Midasin (MIDAS-containing protein).  
 GN MDN1 OR KIA0301.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX Garbarino J.E., Gibbons I.R.;  
 RT "Expression and genomic analysis of midasin, a novel and highly  
 conserved AAA protein distantly related to dynein.";  
 RL BMC Genomics 3:18-18(2002).  
 RN [2]  
 RP SEQUENCE OF 1255-2356 AND 3550-5596 FROM N.A.  
 RC TISSUE=Testis;  
 RA Tracey A.;  
 RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 3550-5596 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kocani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 4:141-150 (1997).  
 CC -1- FUNCTION: May function as a nuclear chaperone and be involved in  
 the assembly/disassembly of macromolecular complexes in the  
 CC nucleus.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: Contains 1 VWPB domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute.. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF039325; AAM77722.1; -  
 DR EMBL; AL096678; CAB86660.1; -  
 DR EMBL; AL096678; CAB86661.1; -  
 DR EMBL; AB002299; BAA20761.1; -  
 DR Genew; HGNC:18302; MDN1.  
 DR GO; GO:0005634; C:nucleus; NAS.  
 DR GO; GO:0016887; F:ATPase activity; NAS.  
 DR GO; GO:0003754; P:chaperone activity; NAS.  
 DR GO; GO:0006461; P:protein complex assembly; NAS.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR002035; VWF\_A.  
 DR Pfam; PF00004; AAA; 1.  
 DR SMART; SM00382; VWA; 7.  
 DR SMART; SM00327; VWA; 1.  
 DR PROSITE; PS50234; VMPA; 1.  
 DR Chaperone; ATP-binding; Repeat: Nuclear protein.  
 FT NP\_BIND 329 336 ATP (POTENTIAL).  
 FT NP\_BIND 677 684 ATP (POTENTIAL).  
 FT NP\_BIND 1084 1091 ATP (POTENTIAL).  
 FT NP\_BIND 1390 1397 ATP (POTENTIAL).  
 FT NP\_BIND 1753 1760 ATP (POTENTIAL).  
 FT NP\_BIND 2066 2073 ATP (POTENTIAL).  
 FT DOMAIN 3566 3573 POLY-GLU.  
 FT DOMAIN 4784 4791 POLY-GLU.  
 FT DOMAIN 5008 5013 POLY-GLU.  
 FT DOMAIN 5182 5187 POLY-GLU.  
 FT DOMAIN 5384 5583 VMPA.  
 FT CONFLICT 2287 2312 RFLSMDFPHGDISRAMRRNGLEIYI -> S (IN REF.  
 FT SEQUENCE 5596 AA; 632802 MW; 586C62616A1P96D4 CRC64;  
 SQ

Query Match 10.1%; Score 86; DB 1; Length 5596;  
 Best Local Similarity 23.4%; Pred. No. 3.5e+02;  
 Matches 30; Conservative 27; Mismatches 63; Indels 8; Gaps 2;

QY 18 LKPRVTVTQNGKMSSTVSEDFILPVYKGELEKQ-----YQDGMETSGFEKGD 71  
 DB 4629 LATHRSTAKLSTLAQVFTLQAKGFCLEPFEMEDSAGGATEFHDEGGGIGEGGMD 4688  
 QY 72 AGVINTLSKDTPIKPKKLEKEENKPTPVSKKNPQVNHQNLNESHKELQRE 131  
 DB 4689 VS--DQIGNEQVEDTFQKQEKEDPDPSKSDIKGDNATKMSDFDGMHDELEQ 4746  
 QY 132 EHSOKSDS 139  
 DB 4747 EDDKSDS 4754

RESULT 24  
 DRS1 YEAST STANDARD; PRT; 752 AA.  
 AC P32852;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Probable ATP-dependent RNA helicase DRS1.  
 GN DRS1 OR YIL008W OR L1345.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93087480; PubMed=1454790;  
 RA Ripmaster T.L., Vaughn G.P., Woolford J.L. Jr.;  
 RT "A putative ATP-dependent RNA helicase involved in Saccharomyces  
 RT cerevisiae ribosome assembly.";

RL Proc. Natl. Acad. Sci. U.S.A. 89:11131-11135(1992).  
 RP  
 RN  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=5288C / FY23;  
 RX MEDLINE=96405918; PubMed=8810043;  
 RA Miosga T., Zimmermann F.K.;  
 RT "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on  
 RT a 43.7 kb fragment of chromosome XII including an open reading frame  
 RT homologous to the human cystic fibrosis transmembrane conductance  
 RT regulator protein CFTR.";  
 RL Yeast 12:693-708(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C / AB972;  
 RX MEDLINE=97313267; PubMed=9169871;  
 RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansgorge W.,  
 RA Bene V., Bruckner M., Deltis H., Dubois E., Duesterhoeft A.,  
 RA Eutlian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,  
 RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,  
 RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,  
 RA Mueller-Auer S., Newkirk U., Obermayer B., Piravandi E., Pohl T.M.,  
 RA Portelle D., Purnelle B., Reckmann S., Rieger M., Rinke M., Rose M.,  
 RA Scharfe M., Scherens B., Scholler P., Schwager C., Schwarz S.,  
 RA Underwood A.P., Urrestazu L.A., Vandemol M., Verhaselt P.,  
 RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambut R., Wedler E.,  
 RA Wedler H., Zimmermann F.K., Zollner A., Hant J., Hohelsel J.D.;  
 RL "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";  
 Nature 387:87-90(1997).  
 CC - FUNCTION: PROBABLE HELICASE INVOLVED IN RIBOSOME ASSEMBLY.  
 CC - SUBCELLULAR LOCATION: Nuclear.  
 CC - SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L00683; AAA4666.1; -  
 DR EMBL; X91488; CAA62783.1; -  
 DR EMBL; Z73113; CAA97452.1; -  
 DR PIR; S64750; S64750.  
 DR HSSP; 058083; LHV8.  
 DR SGD; S0003931; DRS1.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR000629; DEAD\_box.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR Pfam; PF00270; DEAD; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELIC; 1.  
 DR PROSITE; PS00039; DEAD\_ATP\_HELICASE; 1.  
 KW ATP-binding; RNA-binding; Helicase; Nuclear protein.  
 FT DOMAIN 170 190 POLY-GLU.  
 FT NP\_BIND 275 282 ATP (POTENTIAL).  
 FT SITE 365 388 DEAD BOX.  
 FT CONFLICT 1 68 MYVGTKYSNLFVPTISSEDDVPISSDDEKYEAKTT  
 FT SEQUENCE 752 AA; 84843 MW; 60747607A6E5E4A8 CRC64;  
 SQ

Query Match 10.1%; Score 85.5; DB 1; Length 752;  
 Best Local Similarity 21.8%; Pred. No. 43;  
 Matches 38; Conservative 29; Mismatches 64; Indels 43; Gaps 8;

QY 7 IINKDGEVSELPKPRVTVTQNGKMSSTVSEDFILPVYKGELEKGYPD----- 59  
 DB 26 IIDSDDKYEAKK--TTKRRKGR-NGKYSSEGDNDVEDHEDIDAGFKDLDADPTT 81  
 QY 60 ---GMEI--SGFEKQAGVYINLSKDTPIKPKK-----IEEKEENKPTPD 104

Db 82 SNF0CMFLAE6SNKDAEAFV--KKVDLDKIIRKGLVMAHIDSQEEETEKE-K 138  
 QY 105 VSKKKDN-----PQVNHSQLNESHKREKLOREHSHQSKSDSTKVT 144  
 Db 139 VEKENDSDDEELAMDQFGMGAPMANNNDENOSBEEBEEBEEBEEBEEBEEBEMT 192

RESULT 25  
 RGAI\_YEAST STANDARD; PRT; 1007 AA.

AC P39083; P39934;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Rho-type GTPase-activating protein 1  
 GN RGAI OR DBM1 OR THE1 OR YOR127W OR O3290 OR YOR3290W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 NCBI\_TaxID=4932;

OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
 RC STRAIN=S288C;  
 RX MEDLINE=96239492; PubMed=8657111;  
 RA Chen G.-C., Zheng L., Chan C.S.M.;  
 RT "The LIM domain-containing Dbm1 GTPase-activating protein is required  
 RT for normal cellular morphogenesis in Saccharomyces cerevisiae.";  
 RL Mol. Cell. Biol. 16:1376-1390(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / FY1679;  
 RX MEDLINE=97060020; PubMed=8904341;  
 RA Wiemann S., Rechmann S., Benes V., Voss H., Schwager C., Vlcek C.,  
 RA Stegemann J., Zimmermann J., Erfle H., Paces V., Ansoerge W.;  
 RT "Sequencing and analysis of 51 kb on the right arm of chromosome XV  
 RT from Saccharomyces cerevisiae reveals 30 open reading frames.";  
 RL Yeast 12:281-288(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=97344368; PubMed=9200815;  
 RX Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,  
 RA Schwager C., Paces V., Sander C., Ansoerge W.;  
 RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV.";  
 RL Yeast 13:655-672(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=96101594; PubMed=7498791;  
 RA Stevenson B.J., Ferguson B., de Virgili C., Bi E., Pringle J.R.,  
 RA Ammerer G., Sprague G.F. Jr.;  
 RT "Mutation of RGAI, which encodes a putative GTPase-activating protein  
 RT for the polarity-establishment protein Cdc42p, activates the  
 RT pheromone-response pathway in the yeast Saccharomyces cerevisiae.";  
 RL Genes Dev. 9:2949-2963(1995).  
 RN [5]  
 RP SEQUENCE OF 570-639 FROM N.A.  
 RC STRAIN=SNY243;  
 RX MEDLINE=93087574; PubMed=1454852;  
 RA Ramer S.W., Elledge S.J., Davis R.W.;  
 RT "Dominant genetics using a yeast genomic library under the control of  
 RT a strong inducible promoter.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:11589-11593(1992).  
 CC -1- FUNCTION: GTPASE-ACTIVATING PROTEIN (GAP) FOR CDC42 AND/OR RHO1.  
 CC NEGATIVE REGULATOR OF THE PHEROMONE-RESPONSE PATHWAY THROUGH THE  
 CC STE20 PROTEIN KINASE. ACTS AT A STEP BETWEEN THE G-PROTEIN AND THE  
 CC MAP KINASE MODULE. DOMINANT SUPPRESSOR OF BUD EMERGENCE DEFECT  
 CC CAUSED BY DELETION OF IPL2/BEW2. INVOLVED IN THE CONTROL OF  
 CC POLARIZED CELL GROWTH AND PROPER BUD SITE SELECTION.  
 CC -1- SIMILARITY: Contains 2 LIM zinc-binding domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics, and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announce/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC -----  
 CC EMBL; U07421; AAA16875.1; -  
 CC EMBL; X90518; CAA62108.1; -  
 CC EMBL; X94335; CAA64046.1; -  
 CC EMBL; Z75035; CAA99326.1; -  
 CC EMBL; X90950; CAA62445.1; -  
 CC EMBL; L02617; AAA35153.1; -  
 CC PIR; S48535; S48535.  
 CC SGD; S0005653; RGAI.  
 CC GO; GO:0005100; F:Rho GTPase activator activity; IPI.  
 CC GO; GO:0004871; F:signal transducer activity; IPI.  
 CC GO; GO:0007015; P:actin filament organization; IMP.  
 CC GO; GO:0007118; P:apical bud growth; IPI.  
 CC GO; GO:0000283; P:establishment of cell polarity (sensu Sacch. . .); IPI.  
 CC GO; GO:0007125; P:invasive growth; IPI.  
 CC GO; GO:0007119; P:isotropic bud growth; IPI.  
 CC GO; GO:0007124; P:pseudohyphal growth; IPI.  
 CC GO; GO:0000750; P:signal transduction during conjugation with. . .; IGI.  
 CC GO; GO:0007264; P:small GTPase mediated signal transduction; IPI.  
 CC InterPro; IPR001781; LIM.  
 CC InterPro; IPR00198; RhogAP.  
 CC Pfam; PF00412; LIM; 1.  
 CC Pfam; PF00620; RhogAP; 1.  
 CC ProDom; PD000094; LIM; 2.  
 CC SMART; SM00132; LIM; 2.  
 CC SMART; SM00324; RhogAP; 1.  
 CC PROSITE; PS00478; LIM DOMAIN 1; 1.  
 CC PROSITE; PS50238; LIM DOMAIN 2; 2.  
 CC KX GTPase activation; Repeat; LIM domain; Metal-binding; Zinc;  
 CC KW Pheromone response.  
 CC FT DOMAIN 13 66 LIM 1.  
 CC FT DOMAIN 70 122 LIM 2.  
 CC FT DOMAIN 791 1006 RHO-GAP.  
 CC FT VARIANT 866 866 V -> A.  
 CC FT VARIANT 898 898 K -> R.  
 CC FT VARIANT 926 926 S -> G.  
 CC FT MUTAGEN 40 40 C->S; BIPOLAR BUDDING.  
 CC FT MUTAGEN 37 37 C->S; BIPOLAR BUDDING.  
 CC FT MUTAGEN 98 98 C->S; BIPOLAR BUDDING.  
 CC FT MUTAGEN 101 101 C->S; BIPOLAR BUDDING.  
 CC FT CONFLICT 457 457 D -> E (IN REF. 4).  
 CC FT CONFLICT 507 507 T -> P (IN REF. 4).  
 CC SQ SEQUENCE 1007 AA; 112831 MW; C805411B5753791 CRC64;

Query Match 10.1%; Score 85.5; DB 1; Length 1007;  
 Best Local Similarity 21.0%; Pred. No. 59;  
 Matches 34; Conservative 34; Mismatches 69; Indels 25; Gaps 3;

QY 11 DTGSEVSEKPRVYTTIONGKEMSTI-----VSEDPFLLPYVYGELEKGVQPDGWEIS 64  
 Db 349 NTGETSQMDPSLSRRVNLNIVEETALQRPVEVVEKDSRPDLGVOQDAEKYSNN 408

QY 65 GFEQKQDGYVYINLSKOTFIPVFKIEKEEENKPPFDVSKKKDNQVNHSQLNESH 124  
 Db 409 SGKGRKISRSLSRKDLMI-----NLKSRATGKQDSNVKLSPAKSVTSRR 454

QY 125 KEDLQREHSHQSKSDSTKVATVID-----KNNISSTKTTNN 161  
 Db 455 SQDLARDNDSDHTGLDTPNSNSTLDILVNNQKSLNYKRFDTN 496

RESULT 26  
 MAPB\_RAT STANDARD; PRT; 2459 AA.  
 ID MAPB\_RAT  
 AC P15205; O62958; O9ER21; O9QW92;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)



DT 28-FEB-2003 (Rel. 41, last annotation update)  
 DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1  
 light chain LC1].  
 GN MAP1B.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NC NCB1\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE OF 1-142 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Testis;  
 RX MEDLINE=96257242; PubMed=8666295;  
 RA Liu D., Fischer I.;  
 RT "Isolation and sequencing of the 5' end of the rat microtubule-  
 associated protein (MAP1B)-encoding cDNA.";  
 RL Gene 172:307-308(1996).  
 RN [2]  
 RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.  
 RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Gli3 tumor;  
 RX MEDLINE=92347374; PubMed=1639092;  
 RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;  
 RT "Identification of two distinct microtubule binding domains on  
 recombinant rat MAP 1B.";  
 RL Eur. J. Cell Biol. 57:66-74(1992).  
 RN [3]  
 RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.  
 RC TISSUE=spinal cord;  
 RX MEDLINE=90059871; PubMed=2555150;  
 RA Rientz A., Greeningloh G., Hermans-Borgmeyer I., Kirsch J.,  
 RA Litzner U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;  
 RT "Neuraxin, a novel putative structural protein of the rat central  
 nervous system that is immunologically related to microtubule-  
 associated protein 5.";  
 RL EMBO J. 8:2879-2888(1989).  
 RN [4]  
 RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.  
 RX MEDLINE=97405699; PubMed=9260743;  
 RA Ma D., Nothias F., Boyne L.J., Fischer I.;  
 RT "Differential regulation of microtubule-associated protein 1B (MAP1B)  
 in rat CNS and PNS during development.";  
 RL J. Neurosci. Res. 49:319-332(1997).  
 CC -1- FUNCTION: The function of brain MAPs is essentially unknown.  
 CC Phosphorylated MAP1B may play a role in the cytoskeletal changes  
 CC that accompany neurite extension. Possibly MAP1B binds to at least  
 CC two tubulin subunits in the polymer, and this bridging of subunits  
 CC might be involved in nucleating microtubule polymerization and in  
 CC stabilizing microtubules.  
 CC -1- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate  
 CC with MAP1A and MAP1B proteins.  
 CC -1- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,  
 CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,  
 CC heart or muscle.  
 CC -1- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic  
 CC nerve levels are high early in development but decrease during  
 CC postnatal development and are low in adults. In dorsal root  
 CC ganglia levels remain high throughout development.  
 CC -1- INDUCTION: By nerve growth factor.  
 CC -1- DOMAIN: Has a highly basic region with many copies of the sequence  
 CC KKEE and KKEIV, repeated but not at fixed intervals, which is  
 CC responsible for the binding of MAP1B to microtubules.  
 CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated  
 CC from MAP1B by proteolytic processing. It is free to associate with  
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region  
 CC of MAP1B (by similarity).  
 CC -1- PTM: Phosphorylated.  
 CC -1- SIMILARITY: TO MAP1A.  
 CC -1- CAUTION: A C-terminal fragment of this protein (residues 1597 to  
 CC 2459) was originally described as neuraxin in Ref.3.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U52950; AAB17068.1; -  
 CC DR EMBL: X60370; CAC16162.1; -  
 CC DR EMBL: X16623; CAA34620.1; ALT\_SEQ.  
 CC DR PIR: A56577; A56577.  
 CC DR InterPro: IPR00102; MAP1B\_neuraxin.  
 CC DR Pfam: PR00414; MAP1B\_neuraxin: 10.  
 CC DR PROSITE: PS00230; MAP1B\_NEURAXIN; 8.  
 CC FT Microtubules; Repeat; Phosphorylation.  
 CC FT CHAIN ? 2459  
 CC FT REPEAT 1869 1885 MAP1B 1.  
 CC FT REPEAT 1886 1902 MAP1B 2.  
 CC FT REPEAT 1903 1919 MAP1B 3.  
 CC FT REPEAT 1920 1936 MAP1B 4.  
 CC FT REPEAT 1937 1953 MAP1B 5.  
 CC FT REPEAT 1954 1970 MAP1B 6.  
 CC FT REPEAT 1988 2004 MAP1B 7.  
 CC FT REPEAT 2005 2021 MAP1B 8.  
 CC FT REPEAT 2022 2038 MAP1B 9.  
 CC FT REPEAT 2039 2055 MAP1B 10.  
 CC FT DOMAIN 559 1035  
 CC FT DOMAIN 588 786  
 CC FT DOMAIN 2224 2312  
 CC FT CONFLICT 127 127 M -> V (IN REF. 1).  
 CC FT CONFLICT 140 140 T -> S (IN REF. 1).  
 CC FT CONFLICT 2112 2112 R -> K (IN REF. 3).  
 CC FT CONFLICT 2169 2169 L -> I (IN REF. 3).  
 CC SQ SEQUENCE 2459 AA; 269497 MW; 2E3F6872DEB8BA2 CRC64;  
 CC -----  
 CC Query Match 10.1%; Score 85.5; DB 1; Length 2459;  
 CC Best Local Similarity 24.4%; Pred. No. 1.5e+02;  
 CC Matches 40; Conservative 26; Mismatches 53; Indels 45; Gaps 8;  
 CC -----  
 CC QY 4 KEFLNKDGEVSELPKRVTVTQNGKMSSTIVSEEDFLPYKGELEKGFQDGWEI 63  
 CC DB 583 EKVIVKDDPKGVSKSPSTKEVPSKEQS-----PV-KAEVA-----EK 622  
 CC QY 64 SFEFGKQAGVYINISKDTFIKPVK-KLEEKKEENKFTFVSKKQDPQVNSQLNES 122  
 CC DB 623 AATESKP-----KYTKGVKKEIKTKPEKKEE-KRKEVAKED----- 662  
 CC QY 123 HKEDLOREHSQSDSTQVATV--DKNNISSKSTNNKX 164  
 CC DB 663 --KTPKKDEKPKGEAKKEIKKEIKKEKKELKKEVKKETPLK 704  
 CC -----  
 CC RESULT 27  
 CC YBLE SCHPO STANDARD; PRT; 719 AA.  
 CC AC Q10342; Q9UR08;  
 CC DT 01-OCT-1996 (Rel. 34, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Hypothetical protein Cl06.14c in chromosome II.  
 CC GN SPBC106.14C.  
 CC OS Schizosaccharomyces pombe (fission yeast).  
 CC OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC OC Schizosaccharomycetes.  
 CC OX NCB1\_TaxID=4896;  
 CC OX NCB1\_TaxID=4896;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=972;  
 CC RX MEDLINE=2148401; PubMed=11859360;  
 CC Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 CC Sgouros J., Peat N., Hayles J., Baker S., Basham C.M.,  
 CC Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 CC Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 CC Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,





CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: A5000012; AAB95750.1; -.  
DR PIR: S73428; S73428.  
DR InterPro: IPR003760; Bmp.  
DR Pfam: PF02608; Bmp; 1.  
DR PROSITE: PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Hypothetical protein; Lipoprotein; Membrane; Signal;  
KW Complete proteome.  
FT SIGNAL 1 26 POTENTIAL.  
FT CHAIN 27 657 HYPOTHETICAL\_LIPOPROTEIN MG040 HOMOLOG.  
FT LIPID 27 27 N-ACYL DIGLYCERIDE (POTENTIAL).  
SQ SEQUENCE 657 AA; 71671 MW; 6147B1A5673606A9 CRC64;  
  
Query Match 10.0%; Score 84.5; DB 1; Length 657;  
Best Local Similarity 16.7%; Pred. No. 44;  
Matches 42; Conservative 36; Mismatches 78; Indels 95; Gaps 5;  
  
QY 8 LNKDTEGSELKPRVTVTIIONGKEM-----SSTVSEE 41  
DB 343 VLLANVEVARRKHTLIIIGVDSAGELIDNQPAPDQLIKKKIIPSSIKALDAIE 402  
QY 42 DFLPYKGELEKGYOPDG-----WEISGEPEKKDAGY 74  
DB 403 NMLIAIQKSDNNNGYGFYNNIGVTSSVGISEAGYEPLIDPVPMKTTQSGKSMATN 462  
QY 75 VINL-----SKDTIKRVEFKIEKE----- 96  
DB 463 MNLKLSSDDTNTKALKREVSTHKGSGDKGIIGKYSNLLTKSTTTVAQKSWTNN 522  
QY 97 ---EENKPTFDV-SKKDNPQVNHSQLNESHREKEDQREHSQKSDSTQVATVLDKNN 152  
DB 523 SGTEQKKLSEVDTKKKEKSKKTQTSNGDSCQNGKETINDIISKSLTITMTMANNV 582  
QY 153 ISSKSTNNPN 163  
DB 583 MSSKKQSSDN 593  
  
RESULT 30  
ID IGA0\_HAEIN STANDARD; PRT; 1694 AA.  
AC P44969;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAI protease).  
GN IGA OR IGA1 OR HI0990.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxId=727;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=serotype D;  
RA Wright A., Fishman Y., Tai F., Plaut A.G.;  
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.  
RN 12  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / KM20 / ATCC 51907;  
RX MEDLINE=95350630; Pubmed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,

RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd".  
RL Science 269:496-512(1995).  
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A  
CC -1- PRODUCING INTRACT FC AND FAB FRAGMENTS.  
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at  
CC certain Pro-I-xaa bonds in the hinge region. No small molecule  
CC substrates are known.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC  
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE  
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE  
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY  
CC SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X59800; -; NOT ANNOTATED\_CDS.  
DR EMBL: U32779; AAC22651.1; -.  
DR PIR: H64106; H64106.  
DR MEROPS: S06.001; -.  
DR TIGR: H10990; -.  
DR InterPro: IPR006315; Autotransport.  
DR InterPro: IPR005546; Autotransporter.  
DR InterPro: IPR000710; IGA\_S6.  
DR InterPro: IPR004899; Pertactin.  
DR Pfam: PF03797; Autotransporter; 1.  
DR Pfam: PF02395; IGA1; 1.  
DR Pfam: PF03125; Pertactin; 1.  
DR PRINTS: PR00921; IGASERPTASE.  
DR TIGRPFAM: TIGR01414; autotrans\_bar1; 1.  
KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal;  
KW Complete proteome.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.  
FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).  
FT ACT\_SITE 288 288 PROBABLE.  
FT CONFLICT 253 254 EN -> GV (IN REF. 1).  
FT CONFLICT 272 272 G -> A (IN REF. 1).  
FT CONFLICT 464 464 G -> E (IN REF. 1).  
FT CONFLICT 866 866 S -> T (IN REF. 1).  
FT CONFLICT 1036 1036 A -> D (IN REF. 1).  
FT CONFLICT 1074 1074 A -> G (IN REF. 1).  
FT CONFLICT 1421 1421 A -> G (IN REF. 1).  
FT CONFLICT 1545 1545 H -> T (IN REF. 1).  
SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;  
  
Query Match 10.0%; Score 84.5; DB 1; Length 1694;  
Best Local Similarity 26.1%; Pred. No. 1.2e+02;  
Matches 24; Conservative 15; Mismatches 50; Indels 3; Gaps 1;  
  
QY 76 ILNLSQTFKIPVKKLEKEEENKPTFVSKKQNPQVNHSQLNESHREKEDQREHS- 134  
DB 1288 INTGSAITLITFAESDKQETASTEDASQHKANTYADNSVANNSSSDPKSRRSI 1347  
QY 135 --QKSDSTQVATVLDKNNISSKSTNNPNK 164  
DB 1348 SQPQETSAETTAATSTDETTIADNSKSKMNR 1379  
  
RESULT 31  
ID HG2\_HALRO STANDARD; PRT; 510 AA.  
HG2\_HALRO

AC Q02508;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Protein HG2.  
GN HG2.  
OS Halocynthia roretzi (Sea squirt).  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
OC Stolidobranchia; Pyrosidae; Halocynthia.  
OX NCBI\_Taxid=7729;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Oocyte;  
RX MEDLINE=93224498; PubMed=8468323;  
RA Fujiwara S., Kawahara H., Makabe K.W., Satoh N.;  
RT "A complementary DNA for an ascidian embryonic nuclear antigen Hg2  
RT encodes a protein closely related to the amphibian histone-binding  
RT protein N1.";  
RL J. Biochem. 113:189-195 (1993).  
CC -1- FUNCTION: MAY FUNCTION AS A NUCLEOSOME ASSEMBLY FACTOR DURING  
CC RAPID EMBRYONIC CELL DIVISIONS.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- TISSUE SPECIFICITY: EMBRYO AND LARVAE.  
CC -1- DEVELOPMENTAL STAGE: THIS PROTEIN IS DETECTED IN THE NUCLEI  
CC OF ALL CELLS IN EMBRYOS AND LARVAE BUT IS NOT DETECTED IN THE  
CC CELLS OF METAMORPHOSED JUVENILES.  
CC -1- SIMILARITY: TO XENOPUS LAEVIS HISTONE BINDING PROTEIN N1.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL, D13541; BAA02741.1; -.  
DR PIR: JX0254; JX0254.  
DR InterPro: IPR001440; TPR.  
DR Pfam: PF00515; TPR; 2.  
KW Nuclear protein; Developmental protein.  
FT DOMAIN 98 143 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 219 228 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 444 451 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 465 471 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
SQ SEQUENCE 510 AA; 56871 MW; D9961E5953E976FF CRC64;  
  
Query Match 9.9%; Score 84; DB 1; Length 510;  
Best Local Similarity 28.4%; Pred. No. 36;  
Matches 50; Conservative 25; Mismatches 61; Indels 40; Gaps 12;  
  
QY 4 KEFILL-KDTGSEVSELPKPHRTVTITQNGKMSSTVSEEDFLPYKGELEKGYFDGWE 62  
DB 354 KEIIESECKEVEGLKEKLP-----DINSKI--EDVILA--KKQMK--LDG-- 392  
QY 63 ISGF----EGKDAAGVINTSKDTPIK--PVK---KIEEKEEENKPTFVSKKKKNP 112  
DB 393 -SPFQASGESESSGGASTSDKPCSTPIPKVAPTSPVAKDSDPTTLVRRKRPS 451  
QY 113 ----OVNHSQLESREKEDLQREHSQKSDSTKDTATATLDKNNISKSSTNNPNK 164  
DB 452 DEDNQPAESKESKAKAKOEETBEATNGHSAVKDT-DVTDKNGNGHSGKT--PKK 504  
  
RESULT 32  
RBP2\_PLAVB STANDARD; PRT; 1251 AA.  
AC Q00759;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Reticulocyte binding protein 2 (Fragment).  
GN RBP2.

OS Plasmodium vivax (strain Belm).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_Taxid=31273;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92315338; PubMed=1617731;  
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
RT "A reticulocyte-binding protein complex of Plasmodium vivax  
RT merozoites.";  
RL Cell 69:1213-1226 (1992).  
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO  
CC HUMAN RETICULOCYTE CELLS.  
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL, M88098; AAA29744.1; -.  
KW Malaria; Receptor; Membrane.  
FT NON\_TER 1 1  
FT NON\_TER 1 1  
SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;  
  
Query Match 9.9%; Score 84; DB 1; Length 1251;  
Best Local Similarity 23.0%; Pred. No. 96;  
Matches 42; Conservative 37; Mismatches 72; Indels 32; Gaps 9;  
  
QY 4 KEFILLND-TGSEVSELPK--RTVTITQNGKMSST-----VSEEDFLPYKGE 52  
DB 253 REIMMKKOEIKSYSEIKYKDKCTEISNGRGDKLEFEKFPNSENKNNINET 312  
QY 53 EKGYPDGMSEISGF--GKDAAGVINT--SKDTPIKPVKK-----IEEKEEENKPTFD 104  
DB 313 NENINSEQYKLDIEDAKQKSTKVELFKHKEHTTISNFKSEELIGVETSKQKINKAD 372  
QY 105 VSK--KKQNPQV-----NHSQLESREKEDLQREHSQKSDSTKDTATATLD--KN 151  
DB 373 IMKEIERNSSEIQGVKFGQENLKNLNEPHYDNADELDNDKSTNAVLLJETNLESYGH 432  
QY 152 NIS 154  
DB 433 NLS 435  
  
RESULT 33  
FENR\_PEA STANDARD; PRT; 360 AA.  
AC P10933;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Ferredoxin-NADP reductase, leaf isozyme, chloroplast precursor  
DE (EC 1.18.1.2) (FNR).  
GN PETH.  
OS Pisum sativum (Garden pea).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.  
OX NCBI\_Taxid=3888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Little Marvel; TISSUE=leaf;  
RA Newman B.O., Gray J.C.;  
RT "Characterisation of a full-length cDNA clone for pea ferredoxin-NADP+  
RT reductase.";  
RL Plant Mol. Biol. 10:511-520 (1988).  
RN [2]  
SQ SEQUENCE OF 270-360 FROM N.A., AND MUTAGENESIS.

RX MEDLINE=93374905; PubMed=8366077;  
 RA Orellano E.G., Calcaterra N.B., Carrillo N., Ceccarelli E.A.;  
 RT "Probing the role of the carboxyl-terminal region of ferredoxin-NADP+  
 reductase by site-directed mutagenesis and deletion analysis.";   
 RL J. Biol. Chem. 268:19267-19273(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.70 ANGSTROMS).  
 RA MEDLINE=99396739; PubMed=10467097;  
 RA Deng Z., Aliverti A., Zanetti G., Arakaki A.K., Octado J.,  
 RA Orellano E.G., Calcaterra N.B., Ceccarelli E.A., Carrillo N.,  
 RA Karpus P.A.;  
 RT "A productive NADP+ binding mode of ferredoxin-NADP+ reductase  
 revealed by protein engineering and crystallographic studies.";   
 RL Nat. Struct. Biol. 6:847-853(1999).  
 CC -1- FUNCTION: MAY PLAY A KEY ROLE IN REGULATING THE RELATIVE AMOUNTS  
 CC OF CYCLIC AND NON-CYCLIC ELECTRON FLOW TO MEET THE DEMANDS OF THE  
 CC PLANT FOR ATP AND REDUCING POWER.  
 CC -1- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized  
 CC ferredoxin + NADPH.  
 CC -1- COFACTOR: FAD.  
 CC -1- PATHWAY: FINAL STEP IN LINEAR PHOTOSYNTHETIC ELECTRON TRANSPORT  
 CC CHAIN; IT HAS ALSO BEEN IMPLICATED IN CYCLIC ELECTRON FLOW AROUND  
 CC PHOTOSYSTEM I, ITS ROLE BEING TO RETURN ELECTRONS FROM FERREDOXIN  
 CC TO THE CYTOCHROME B-F COMPLEX.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST; STROMAL SIDE OF THE THYLAKOID  
 CC MEMBRANE IN THE VICINITY OF THE MEMBRANE.  
 CC -1- MISCELLANEOUS: FNR IS PROBABLY ATTACHED TO THE MEMBRANE BY A  
 CC SPECIFIC BINDING PROTEIN.  
 CC -1- SIMILARITY: WITH OTHER SPECIES FNR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X12446; CAA30978.1; -;  
 CC EMBL; L15565; AAB59349.1; -;  
 CC EMBL; L15567; AAB59303.1; -;  
 CC EMBL; L15569; AAB59304.1; -;  
 CC F1R; S04030; S04030.  
 CC PDB; 1QF2; 01-SEP-99.  
 CC PDB; 1QFY; 01-SEP-99.  
 CC PDB; 1QGO; 01-SEP-99.  
 CC PDB; 1QGN; 01-SEP-99.  
 CC InterPro: IPR001709; FPN\_cyt\_reductase.  
 CC InterPro: IPR001433; Oxired\_FAD/NAD(P).  
 CC Pfam: PF00175; NAD\_binding\_1; 1.  
 CC PRINTS: PR00371; FPNCR.  
 CC Oxidoreductase; Flavoprotein; NADP; FAD; Chloroplast; Transit peptide;  
 CC Electron transport; Photosynthesis; Thylakoid; Membrane;  
 CC Multigene family; 3D-structure.  
 CC TRANSIT 1 52 CHLOROPLAST.  
 CC CHAIN 53 360 FERREDOXIN-NADP REDUCTASE, LEAF ISOZYME.  
 CC DISULFID 178 183 BY SIMILARITY.  
 CC NP BIND 212 230 NADP (RIBOSE PART) (BY SIMILARITY).  
 CC MUTAGEN 360 360 Y-W,F,S.  
 CC STRAND 68 68  
 CC TURN 71 72  
 CC STRAND 78 78  
 CC TURN 80 81  
 CC STRAND 84 93  
 CC TURN 97 98  
 CC STRAND 103 109  
 CC TURN 111 112  
 CC STRAND 118 119  
 CC TURN 121 125  
 CC STRAND 129 129  
 CC TURN 131 132

FT STRAND 135 135  
 FT STRAND 139 143  
 FT TURN 147 148  
 FT TURN 150 151  
 FT STRAND 156 162  
 FT STRAND 165 167  
 FT TURN 169 170  
 FT STRAND 173 175  
 FT STRAND 177 184  
 FT TURN 187 188  
 FT STRAND 190 197  
 FT TURN 200 201  
 FT STRAND 204 204  
 FT TURN 207 208  
 FT STRAND 210 216  
 FT HELIX 217 220  
 FT HELIX 221 231  
 FT STRAND 232 232  
 FT STRAND 243 250  
 FT HELIX 253 255  
 FT TURN 257 258  
 FT HELIX 259 268  
 FT TURN 270 272  
 FT STRAND 273 279  
 FT TURN 280 282  
 FT STRAND 284 284  
 FT TURN 286 287  
 FT STRAND 290 290  
 FT HELIX 293 297  
 FT HELIX 298 300  
 FT HELIX 301 309  
 FT TURN 311 312  
 FT STRAND 313 319  
 FT TURN 321 322  
 FT HELIX 323 337  
 FT TURN 338 339  
 FT HELIX 342 351  
 FT TURN 352 353  
 FT STRAND 355 360  
 SQ SEQUENCE 360 AA; 40194 MW; 7F1C10DEBBA7B24 CRC64;  
 Query Match 9.8%; Score 83.5; DB 1; Length 360;  
 Best local Similarity 24.7%; Pred. No. 27;  
 Matches 38; Conservative 19; Mismatches 50; Indels 47; Gaps 7;  
 QY 3 VKEPIINKDTGEV-----SELKP-HRVTYTIQNGKEM-----SRTI 37  
 Db 161 VKRLVYTNDAAGVVGVCNSFLCDLPGSEVKITGPVGKEMMPXDPNATVIMLTGTGTGI 220  
 QY 38 VEEEDPILPVYKGELEKGYQFDG--WEISGFEGKKDAGVIVNLSKDTFIKPVFKKIEBK 95  
 Db 221 ADFRSFLTMWF-FEKHEDYQFNGLAMFLG-----VPTSSSLYKEFEKWKKEKA 269  
 QY 96 EENKPTFPVSKKKONPVNHSQNLNESHREKDLQ 129  
 Db 270 PENFRIDFAVRSR-----QVNDKGEKWTIQ 294  
 RESULT 34  
 ID U2R1 HUMAN STANDARD; PRT; 479 AA.  
 AC Q15655; Q13570;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE U2 small nuclear ribonucleoprotein auxiliary factor 35 kDa subunit  
 DE related-protein 1.  
 GN UZAF1-RS1 OR UZAF1RS1 OR UZAFBPL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RC TISSUE=Brain:
RX MEDLINE=96163878; PubMed=8586425;
RA Kitagawa K., Wang X., Hatada I., Yamaoka T., Nojima H.,
RA Itazawa J., Abe T., Mitsuya K., Oshimura M., Murata A., Monden M.,
RA Mukai T.,
RA "Isolation and mapping of human homologues of an imprinted mouse gene
RT U2af1-rsl.",
RL Genomics 30:257-263(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96212931; PubMed=863064;
RA Parrall R.S., Shibata H., Brozowska A., Yoshino K., Okuda K.,
RA DeJong P.J., Plase C., Chapman V.M., Hayashizaki Y., Held W.A.;
RT "Absence of imprinting in U2AFBPL, a human homologue of the imprinted
RL mouse gene U2af1b-rs.",
CC Biochem. Biophys. Res. Commun. 222:171-177(1996).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -1- SIMILARITY: Contains 2 C3H1-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: D49676; BAA08532.1; -.
DR EMBL: U51224; AAB98669.1; -.
DR MIM: 601079; -.
DR GO: GO:0005634; C:nucleus; NAS.
DR GO: GO:0003723; F:RNA binding activity; NAS.
DR InterPro: IPR000504; RNA_rec_mot.
DR InterPro: IPR000571; ZnF_CCCH.
DR Pfam: PF00076; rrm; 1.
DR Pfam: PF00642; ZF-CCCH; 2.
DR SMART: SM00360; RRM; 1.
DR SMART: SM00356; ZnF_C3H1; 2.
DR PROSITE: PS00102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
DR Nuclear protein; RNA-binding; Ribonucleoprotein; Zinc-finger; Repeat.
FT DOMAIN 203 309 RNA-BINDING (RRM).
SQ SEQUENCE 479 AA; 57643 MW; 96F326694BD4E7C0 CRC64;
Query Match 9 8%; Score 83.5; DB 1; Length 479;
Best local Similarity 21.7%; Pred.No.37;
Matches 26; Conservative 25; Mismatches 40; Indels 29; Gaps 4
QY 70 KDAGVINSKDTFIPVKIIEKK-----EEENKPTFVSKKKDNQ 113
DB 44 RDSGLSGEEBETFLIE--QGLEEKLERERRLHEWLRQGAQGEFRFKKEKEEA 101
QY 114 VN-----HSQINSHRKEDLQREHSQKSDSTQVATVLD--KNNISKSTNNP 162
DB 102 KKWLEERQKLEKEQWKEQQRKEREBEBOQKQKEKEKEAVQKMLDQAEIDLNSTTWQNP 161
RESULT 35
AC GARP_PLAUF STANDARD; PRT; 678 AA.
ID GARP_PLAUF
PI 3816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutamic acid-rich protein precursor.
GN GARP.
OS Plasmodyon faiciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_Taxid=5837;
RP SEQUENCE FROM N.A.

```

RA MEDLINE=99040048; PubMed=2903445; Silva A., Anders R.F.,  
RX Triglia T., Stahl H.-D., Crewther P.E.,  
RA Kemp D.J.,  
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic  
RT acid-rich protein (GARP).";  
RL Mol. Biochem. Parasitol. 31:199-202(1988).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC  
CC EMBL; J03998; AAA29605.1; -  
DR PIR; A54514; A54514.  
KW Repeat; Malaria; Antigen; Signal.  
FT SIGNAL 1 25  
FT CHAIN 26 678 GLUTAMIC ACID-RICH PROTEIN.  
FT DOMAIN 120 164 15 X 3 AA TANDEM REPEATS OF K-K-X.  
FT FT 372 416 9 X APPROXIMATE TANDEM REPEATS.  
FT FT 417 441 5 X APPROXIMATE TANDEM REPEATS.  
FT DOMAIN 576 604 POLY-GLU.  
FT DOMAIN 605 653 7 X APPROXIMATE TANDEM REPEATS.  
FT FT 654 663 POLY-GLU.  
SQ SEQUENCE 678 AA; 80551 MW; 2A8F8560649E9A9E CRC64;  
Query Match 9.8%; Score 83.5; DB 1; Length 678;  
Best Local Similarity 24.6%; Pred. No. 54;  
Matches 44; Conservative 33; Mismatches 65; Indels 37; Gaps 10;  
OY 7 ILAND---TGEV---SELKPHRVTVTIQNGKENSSTIVSE---DFILFVYKGELEKG 55  
DB 37 ILNKSFSISITRLNLTNETLEKKN-----DONSKESTLKEKDEKQDVPPTNSNDLNKA 90  
OY 56 YQPGWMEISGEGGKKDAGVYINLSKDFIKFVFKKIEEK-----EBENKPTDVSKKK 109  
DB 91 HNNN--EIS--SSTDPTNIIIVNDKDNENSVDKKKKKKKKKKKKKKKKKKKKKKKKKK 145  
OY 110 DNPQVNSQLNESHREKEDLQREHSQ-----KSDSTQDVTATVLDKNNISSK--STTN 161  
DB 146 DKKKKKKK--EKKKKKKKKKKENSEVMSLYKTQHKPKKATTEGENSELDSEMSSEINN 202  
RESULT 36  
HS9A\_HUMAN STANDARD; PRT; 731 AA.  
ID\_HS9A\_HUMAN  
AC P07900; G9BVO5; 01-AUG-1988 (Rel. 08, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Heat shock protein HSP 90-alpha (HSP 86).  
GN HSPCA OR HSPCL OR HSP90A.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP TISSUE=PerIPHERAL blood lymphocytes;  
RX Soeda E., Yokoyama K., Yamazaki M., Akaogi K., Miwa T., Imai T.;  
RT "Nucleotide sequence of a full-length cDNA for 90 kDa heat-shock  
RT protein from human peripheral blood lymphocytes.";  
RL Nucleic Acids Res. 17:7108-7108(1989).  
RN [2]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=91242090; PubMed=1368637;  
RX Yamazaki M., Taahito H., Yokoyama K., Soeda E.;  
RT "Molecular cloning of cDNA encoding a human heat-shock protein whose  
RT expression is induced by adenovirus type 12 E1A in HeLa cells.";  
RL Agric. Biol. Chem. 54:3163-3170(1990).

[3]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE-Placenta;  
 RX MEDLINE=89343979; PubMed=2527334;  
 RA Hickey E., Brandon S.B., Smale G., Lloyd D., Weber L.A.;  
 RT "Sequence and regulation of a gene encoding a human 89-kilodalton  
 RT heat shock protein.";  
 RL Mol. Cell. Biol. 9:265-2626(1989).  
 [4]  
 RN SEQUENCE OF 1-311 FROM N.A.  
 RX MEDLINE=89056312; PubMed=2445630;  
 RA Hoffmann T., Hovemann B.;  
 RT "Cloning and nucleotide sequence of the murine hsp84 cDNA and  
 RT chromosome assignment of related sequences.";  
 RL Gene 56:29-40(1987).  
 [5]  
 RN SEQUENCE OF 1-311 FROM N.A.  
 RX MEDLINE=90076956; PubMed=2591742;  
 RA Walter T., Drabant B., Krebs H., Tomalak M., Heiss S.,  
 RA Benecke B.J.;  
 RT "Cloning and analysis of a human 86-kDa heat-shock-protein-encoding  
 RT gene.";  
 RL Gene 83:105-115(1989).  
 [6]  
 RN SEQUENCE OF 184-731 FROM N.A.  
 RC TISSUE-Placenta;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Mansina K., Farmer A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [7]  
 RN SEQUENCE OF 538-731 FROM N.A.  
 RC TISSUE-Heart;  
 RA Tanaka M., Tanaka T., Mitsui Y., Yamamoto M., Wood J.N.;  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 [8]  
 RN SEQUENCE OF 1-20, AND PHOSPHORYLATION.  
 RX MEDLINE=89123325; PubMed=2492519;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "Two human 90-kDa heat shock proteins are phosphorylated in vivo at  
 RT conserved serines that are phosphorylated in vitro by casein kinase  
 RT II.";  
 RL J. Biol. Chem. 264:2431-2437(1989).  
 [9]  
 RN PHOSPHORYLATION BY DS-DNA KINASE.  
 RX MEDLINE=9008887; PubMed=2507541;  
 RA Lees-Miller S., Anderson C.W.;  
 RT "The human double-stranded DNA-activated protein kinase phosphorylates  
 RT the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal  
 RT threonine residues.";  
 RL J. Biol. Chem. 264:17275-17280(1989).  
 [10]  
 RN INTERACTION WITH OM34.  
 RX MEDLINE=98324997; PubMed=9660753;  
 RA Young J.C., Obermann W.M., Hartl F.U.;  
 RT "Specific binding of tetratricopeptide repeat proteins to the

RT C-terminal 12-kDa domain of hsp90.";  
 RL J. Biol. Chem. 273:18007-18010(1998).  
 [11]  
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 10-222.  
 RP MEDLINE=97262065; PubMed=9108479;  
 RX Stebbins C.E., Russo A.A., Schneider C., Rosen N., Hartl F.U.,  
 RA Pavletich N.P.;  
 RT "Crystal structure of an Hsp90-geldanamycin complex: targeting of a  
 RT protein chaperone by an antitumor agent.";  
 RL Cell 89:239-250(1997).  
 [12]  
 RN X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 10-222.  
 RP MEDLINE=99034582; PubMed=9817749;  
 RA Obermann W.M., Sandermann H., Russo A.A., Pavletich N.P., Hartl F.U.;  
 RT "In vivo function of Hsp90 is dependent on ATP binding and ATP  
 RT hydrolysis.";  
 RL Cell Biol. 143:901-910(1998).  
 CC -1- FUNCTION: Molecular chaperone. Has ATPase activity  
 CC (By similarity).  
 CC -1- SUBUNIT: Homodimer. Interacts with OM34.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X15183; CAA33259.1; -  
 DR EMBL: X07270; CAA30255.1; -  
 DR EMBL: M27024; AAA31394.1; -  
 DR EMBL: M30626; AAA36023.1; -  
 DR EMBL: BC000987; AAH00987.1; -  
 DR EMBL: D87666; BAA13430.1; -  
 DR EMBL: D87666; BAA13431.1; -  
 DR PIR: A32319; HHHU86.  
 DR PDB: 1YER; 22-APR-98.  
 DR PDB: 1YES; 22-APR-98.  
 DR PDB: 1YET; 22-APR-98.  
 DR PDB: 1BYO; 28-OCT-98.  
 DR TRASNAC: T00992; -  
 DR GeneW: HGNC:5253; HSPCA.  
 DR MIM: 140571; -  
 DR GO: GO:0005737; C:cytoplasm; NAS.  
 DR GO: GO:0003773; F:heat shock protein activity; NAS.  
 DR InterPro: IPR003594; AtPbind\_Arpase.  
 DR InterPro: IPR001404; Hsp90.  
 DR Pfam: PF02518; Hspase\_C1.  
 DR Pfam: PF00183; HSP90\_1.  
 DR PRINTS: PR00775; HEATSHOCK90.  
 DR SMART: SM00387; HATPase\_C1.  
 DR PROSITE: PS00298; HSP90\_1.  
 DR Chaperone; ATP-binding; Heat shock; Phosphorylation; 3D-structure.  
 FT INIT MET 0 0  
 FT MOD RES 4 4  
 FT MOD RES 6 6  
 FT MOD RES 230 230  
 FT MOD RES 262 262  
 FT CONFLICT 62 62  
 FT STRAND 17 20  
 FT HELIX 23 34  
 FT TURN 40 41  
 FT HELIX 42 64  
 FT HELIX 66 69  
 FT TURN 70 71  
 FT STRAND 77 82  
 FT TURN 83 86  
 FT STRAND 87 92  
 FT HELIX 99 103  
 FT TURN 104 107  
 T -> S (IN REF. 3, 4 AND 5).

```

FT HELIX 110 122
FT TURN 123 124
FT HELIX 127 133
FT TURN 134 133
FT HELIX 136 142
FT TURN 143 143
FT STRAND 144 152
FT TURN 154 155
FT STRAND 158 163
FT TURN 166 167
FT STRAND 168 173
FT STRAND 182 189
FT HELIX 191 197
FT HELIX 199 209
FT TURN 211 212
FT STRAND 217 219
SQ SEQUENCE 731 AA; 84542 MW; 1249ABCFCE06297C CRC64;

Query Match 9.8%; Score 83.5; DB 1; Length 731;
Best Local Similarity 23.6%; Pred. No. 58;
Matches 39; Conservative 28; Mismatches 59; Indels 39; Gaps 7;

OY 6 FILNKDGEVSEELKRVVTITQNGKEMSSITVSEDEFILPYKGELEKGYQFDCWEISG 65
DB 169 FTVRTDICE-----PMGRGTKVILHLKEDQTEYLEERRI-----KEIVKHSGFIDGPIITL 219
OY 66 F-EGKKDAGYVNIISKDTFIKVFKKIEKKEENKPTFDVSKKKNQPNVNSHSHR 124
DB 220 FVEKERDK-----EVSDD-----EAEKEDKEEKEKEKESKSDKPEI----- 257
OY 125 KEDLQREHSHQKSDSTKVATATVLDKN-----NISKSTTNPN 163
DB 258 -EDVGSDEEEKKDDKDKKKKKIKKKYIDQELNKTPIWTNPD 301

RESULT 37
ARS2_DROME
AC 09V9K7;
ID ARS2_DROME STANDARD; PRT; 943 AA.
AC 09V9K7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Arginine-resistance protein 2 homolog.
GN CG7843.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Abmayant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Caceres A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downe M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster K., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

```

```

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Martel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RA Science 287:2185-2195 (2000).
RL [2]
RN REVISIONS, AND ALTERNATIVE SPLICING.
RP STRAIN=Berkeley;
RX MEDLINE=22426069; PubMed=12537572;
RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Bernier B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter U., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RA "Annotation of the Drosophila melanogaster euchromatic genome: a
RA systematic review."
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22 (2002).
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q9V9K7-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q9V9K7-2; Sequence=VSP 000327;
CC -I- SIMILARITY: BELONGS TO THE ARS2 FAMILY.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL, AE003784; AAM68343.1; -
CC DR EMBL, AE003784; AAM68345.1; -
CC DR Flybase; FBgn003062; CG7843.
CC DR Pfam; PF04959; ARS2; 1.
CC FT Hypothetical protein; Alternative splicing.
CC FT VARSPLIC 47 50 Missing (in isoform short).
CC FT /ftid=VSP 000327
SQ SEQUENCE 943 AA; 107221 MW; 0C1AF09E02EBA0B CRC64;

Query Match 9.8%; Score 83.5; DB 1; Length 943;
Best Local Similarity 25.0%; Pred. No. 77;
Matches 36; Conservative 23; Mismatches 68; Indels 17; Gaps 4;

OY 14 EYSELKP-----HRVVTITQNGKEMSSITVSEDEFILPYKGELEKGYQFDCWEISG 61
DB 272 KYLDEKDPVYVEKAKQOMQSVKEVKTINSPKEKMSADPVSTQRPVPPVNSDGENW 331
OY 62 EISGEGKADAGYVNIISKDTFIKVFKKIEKKEENKPTFDVSKKKNQPNVNSHSHR 118
DB 332 DDDDAENAPPKKELAEKSDSKSDPKQKLNKKTKKKRASSDDSSSSSSSSSSSSSDEEK 391
OY 119 LNESHKEDLQREHSHQKSDSTKD 142
DB 392 LKEKYVDVGLRAE--QKTEAKND 413

```



RESULT 38  
 IF31 HUMAN STANDARD; PRT; 258 AA.  
 ID IF31\_HUMAN  
 AC 075822; Q9BUD2; Q9H8Q2;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Eukaryotic translation initiation factor 3 subunit 1 (eIF-3 alpha)  
 DE (eIF3 p35) (eIF3j).  
 GN EIF3J1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NC NCBI\_TaxID=9606;  
 OK  
 RN (1)  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=99041954; PubMed=9822659;  
 RA Block K.L., Vornlocher H.-P., Hershey J.W.B.;  
 RT "Characterization of cDNAs encoding the p44 and p35 subunits of human  
 translation initiation factor eIF3.";  
 RL J. Biol. Chem. 273:31901-31908 (1998).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ovarian carcinoma;  
 RA Isegaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Satto K.,  
 RA Yamamoto J., Makanatsu A., Nakamura Y., Nagahara K., Masuo Y.,  
 RA Niinomiya K., Iwayanagi T.;  
 RT "NEO human cDNA sequencing project.";  
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strassberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udell T.B., Toshlyuk S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Maira M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 CC -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF  
 CC METIONYL-TRNAI AND MRNA.  
 CC -1- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 12 DIFFERENT SUBUNITS (BY  
 CC similarity).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL, U97670; AAC78729.1; -.  
 CC DR EMBL, AK023386; BAB14555.1; -.

DR EMBL, BC002719; AA02719.1; -.  
 DR Genem; HGN:3270; EIF3S1.  
 DR CK; 075822; -.  
 DR MIM; 603910; -.  
 DR GO; GO:0005852; Eukaryotic translation initiation factor 3 . . .; TAS.  
 DR GO; GO:0003743; Translation initiation factor activity; TAS.  
 DR GO; GO:0004466; P-regulation of translational initiation; TAS.  
 KW Initiation factor; Protein biosynthesis.  
 FT DOMAIN 2  
 FT POLY-ALA.  
 FT POLY-GLY.  
 FT POLY-ASP.  
 FT POLY-LYS.  
 FT DOMAIN 218 224  
 FT E -> G (IN REF. 1).  
 FT CONFLICT 141 141 A -> T (IN REF. 2).  
 SQ SEQUENCE 258 AA; 29062 MW; 8362423542445AA CRC64;  
 Query Match 9.8%; Score 83; DB 1; Length 258;  
 Best Local Similarity 25.4%; Pred. No. 21;  
 Matches 30; Conservative 25; Mismatches 51; Indels 12; Gaps 4;  
 QY 48 YGELFKGYQFPGWETISGFEKKDAGY--VINTSKDTIKPVFKIEKKEKPTDV 105  
 DB 39 WEGBDEDEDVKDMDDDDEKKEBAEVKPKISK--KTKAEKIKERQCKROBEI 95  
 QY 106 SKKKNPQ----VNHSQLNESHREKEDOREHSGQKSDTKDVTATV--LDKNISSK 156  
 DB 96 KKRLEPPEPKVLTPEEQDLKRLKKQESDLELAETGVNNAVYGIAMPSSR 153  
 RESULT 39  
 TONB HAEIN  
 ID TONB\_HAEIN STANDARD; PRT; 270 AA.  
 AC P42872;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE TonB protein.  
 GN TONB OR H10251.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 OC NCBI\_TaxID=727;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ETH1 IN106;  
 RX MEDLINE=94245357; PubMed=8188372;  
 RA Jantosik G.P., Sanders J.D., Cope L.D., Muller-Berhard U.,  
 RA Hansen E.J.;  
 RT "A functional tonB gene is required for both utilization of heme and  
 RT virulence expression by Haemophilus influenzae type b.";  
 RL Infect. Immun. 62:2470-2477 (1994).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirtness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Wertheck J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Goeyne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weisman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Usterbach T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512 (1995).  
 CC -1- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT  
 CC CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO  
 CC THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO  
 CC TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-  
 CC REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE  
 CC RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER

MEMBRANE PROTEINS. REQUIRED FOR HEME UTILIZATION AND VIRULENCE.  
 -1- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE PERIPLASM (BY SIMILARITY).  
 -1- SIMILARITY: BELONGS TO THE TONB FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; U04996; AAA60460.1; -  
 CC EMBL; U32711; AAC21917.1; -  
 CC PIR; F64057; F64057.  
 CC TIGR; H10251; -  
 DR InterPro; IPR003538; TonB\_C.  
 DR InterPro; IPR006260; TonB\_C.  
 DR Pfam; PF03544; TonB; 1.  
 DR PRINTS; PR01374; TONBPROTEIN.  
 DR TIGRFAMs; TIGR01352; tonb Cterm; 1.  
 KW Transport; Protein transport; Inner membrane; Periplasmic; Transmembrane; Signal-anchor; Repeat; Virulence; Complete proteome.  
 FT DOMAIN 1 7  
 FT TRANSMEM 8 27  
 FT DOMAIN 28 270  
 FT DOMAIN 70 77  
 FT DOMAIN 58 90  
 FT DOMAIN 98 143  
 FT VARIANT 17 17  
 FT VARIANT 31 31  
 FT VARIANT 66 66  
 FT VARIANT 76 84  
 FT VARIANT 116 116  
 FT VARIANT 120 120  
 FT VARIANT 134 134  
 FT VARIANT 140 140  
 FT VARIANT 146 146  
 FT VARIANT 226 226  
 FT VARIANT 226 226  
 FT SEQUENCE 270 AA; 29193 MW; 4CD54B46F7D10A0 CRC64;  
 Query Match 9.8%; Score 83; DB 1; Length 270;  
 Best Local Similarity 32.1%; Pred. No. 22;  
 Matches 26; Conservative 16; Mismatches 33; Indels 6; Gaps 3;  
 QY 80 KDTFKPVFKLIEKKKEENKPTFDVSKKKDNPVNHSQLNESHKREDLQREHGXKSDS 139  
 DB 93 EDPFKPEPKKLEPEKPEKPKK--GKPKGKPK--NPKKVEYKPKKKPTNKE--LPKGD 146  
 QY 140 TKDVTATVLDKNNISKSTTN 160  
 DB 147 NIDSSANVNDKASTTSANSN 167  
 RESULT 40  
 ADVS\_HUMAN STANDARD; PRT; 715 AA.  
 ID ADVS\_HUMAN O8WU97;  
 AC 09Y6U3; O8WU97;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-EB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Adseverin (Scinderin).  
 GN SCIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid:9606;  
 RN [1].  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=skin;  
 RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Pelngold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Malahy S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield J.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE OF 1-527 FROM N.A.  
 RA Kalicki J., Smith-Craig R.,  
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 248-715 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arica M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.,  
 RT "NEO human cDNA sequencing project.";  
 RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: Ca(2+)-dependent actin filament severing protein that is presumed to have a regulatory function in exocytosis by affecting the organization of the microfilament network underneath the plasma membrane. In vitro, also has bared end capping and nucleating activities in the presence of Ca(2+).  
 CC -1- SIMILARITY: BELONGS TO THE VILLIN/GELSOLIN FAMILY.  
 CC -1- SIMILARITY: Contains 6 gelsolin-like repeats.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; BC021090; AAH21090.1; -  
 CC EMBL; AC005281; AAD15423.1; -  
 CC EMBL; AK027778; BAB5361.1; -  
 CC HSSP; P02640; 2VTL.  
 DR InterPro; IPR001974; Gelsolin.  
 DR Pfam; PF00626; Gelsolin; 6.  
 DR PRINTS; PR00597; GELSOLIN.  
 DR SMART; SM00262; GEL; 6.  
 KW Cytoskeleton; Actin-binding; Repeat; Calcium; Capping protein. ACTIN-SEVERING (POTENTIAL).  
 FT DOMAIN 1 363  
 FT REPEAT 27 76  
 FT REPEAT 148 188  
 FT REPEAT 265 307  
 FT REPEAT 398 451  
 FT REPEAT 523 564  
 FT REPEAT 626 668  
 FT SITE 112 119  
 FT SITE 138 146  
 FT SITE 61 61  
 FT CONFLICT 715 AA; 80508 MW; 45FBF42CB8FDD80 CRC64;  
 SO SEQUENCE

Query Match 9.8%; Score 83; DB 1; Length 715;  
Best Local Similarity 23.1%; Pred. No. 62;  
Matches 34; Conservative 30; Mismatches 47; Indels 36; Gaps 5;

QY	22	RYVTITIQNGKEMSSITVSEEDFILPYKGELEKGYQFDGWEISGFEGK-----KDAGYV	75
Db	262	RYTVVAEENPFSMAMLLSEECFILD--HGAAQIFVWKGKDANPQERKAAMKTAEEFLQO	319
QY	76	INLSKDTFIK-----PVFKK-----IPEKKEENKPTFDVSKKK	109
Db	320	MNYSKNTQIQVLEPGGETPIFKQFQKMDKQSDSGKVTYEKVAQIKQIIPDASKLH	379
QY	110	DNPQV--NHSOLNESHKEDLOREHS	134
Db	380	SSPQMAAQHNMVDDSGKVEIRVENN	406

Search completed: February 10, 2004, 10:49:55  
Job time : 9.62413 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 8.09802 Seconds  
(without alignments)  
1011.574 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897

Sequence: 1 KIVKDFPARNITVKEFLNK.....ATVLDKNNISSKSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100	11.1	1875	1	MLP1_YEAST
2	99	11.0	1398	1	MLP1_YEAST
3	97.5	10.9	1345	1	YH00_YEAST
4	97.5	10.9	2468	1	MABP_HUMAN
5	96	10.7	893	1	GYRA_STAE
6	95	10.6	348	1	CYLD_HUMAN
7	95	10.6	1202	1	DPMO_ASCIM
8	93.5	10.4	649	1	HS70_PYRSA
9	93	10.4	715	1	ADSV_HUMAN
10	93	10.4	2464	1	MABP_MOUSE
11	92.5	10.3	443	1	DACA_BACSU
12	92.5	10.3	1702	1	IGAZ_HAEIN
13	92.5	10.3	1888	1	YDRT_SCHPO
14	92	10.3	406	1	NSH1_MOUSE
15	92	10.3	1332	1	SPT7_YEAST
16	90	10.0	688	1	LIP_STAE
17	90	10.0	778	1	YFRK_YEAST
18	89.5	10.0	433	1	TIG_STEAM
19	89.5	10.0	1066	1	PIKI_YEAST
20	89.5	10.0	1694	1	IGAO_HAEIN
21	88	9.8	853	1	YCGI_YEAST
22	87.5	9.8	360	1	FENR_PEA
23	87.5	9.8	655	1	SKM1_YEAST
24	87.5	9.8	678	1	GARP_PLAFR
25	87.5	9.8	879	1	RA50_PYRHO
26	87	9.7	6632	1	UN69_CAEEL
27	86.5	9.6	1006	1	RA11_YEAST
28	86.5	9.6	1251	1	RBP2_PLAVB
29	86	9.6	200	1	SYST_LYCES
30	86	9.6	443	1	GLNA_PYRKO
31	86	9.6	914	1	BPA_BACSU
32	86	9.6	5596	1	MDN1_HUMAN
33	85.5	9.5	752	1	DRS1_YEAST

34	85.5	9.5	1007	1	RGAL_YEAST
35	85.5	9.5	2459	1	MABP_RAT
36	85	9.5	365	1	FENR_MESCR
37	85	9.5	472	1	6PGD_LACIA
38	85	9.5	657	1	Y040_MYCPN
39	85	9.5	719	1	YBLE_SCHPO
40	85	9.5	720	1	IF2_STAE
41	85	9.5	895	1	RA50_THEVO
42	84.5	9.4	950	1	IF2_LACIC
43	84	9.4	510	1	HVZ2_HALFO
44	84	9.4	796	1	YFC3_YEAST
45	83.5	9.3	363	1	FENR_VICPA

## ALIGNMENTS

RESULT 1  
ID MLP1\_YEAST STANDARD; PRT: 1875 AA.  
AC Q02455;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin-like protein MLP1.  
GN MLP1 OR YKR095W OR YKR415.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288c;  
RA MEDLINE=93247549; PubMed=8483450;  
RA Koeilling R., Nguyen T., Chen E.Y., Botstein D.;  
RT "A new yeast gene with a myosin-like heptad repeat structure."; Mol. Genet. 237:359-369(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=94205265; PubMed=8154186;  
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G., Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;  
RT "The complete sequence of a 15,820 bp segment of Saccharomyces cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three new open reading frames.";  
RL Year 9:1349-1354(1993).  
CC -!- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA REPAIR.  
CC -!- SIMILARITY: SOME, TO THE TPR ONCOGENE.  
CC -!- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MLP1".  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by, and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, L01992; AAA34783.1; -;  
DR EMBL, X73541; CAAS1948.1; -;  
DR EMBL, Z28320; CAAB2174.1; -;  
DR PIR, S38173; S38173.  
DR SGD, S0001803; MLP1.  
DR GO, GO:0005635; C:nuclear membrane; IDA.  
DR GO, GO:0005654; C:nucleoplasm; IDA.  
DR GO, GO:0006606; P:protein-nucleus import; IDA.  
KW Coiled coil; DNA repair.  
FT DOMAIN 69 487 COILED COIL (POTENTIAL).  
FT DOMAIN 531 1678 COILED COIL (POTENTIAL).  
FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).  
FT CONFLICT 301 301 R -> A (IN REF. 1).  
SQ SEQUENCE 1875 AA; 218455 MW; 683A0D34C906667 CRC64;

```

Query Match      11.1%; Score 100; DB 1; Length 1875;
Best Local Similarity 24.4%; Pred. No.14;
Matches 50; Conservative 38; Mismatches 71; Indels 46; Gaps 10;

QY 4 VKDPARTTVEKILNDIGVSEELKPHRVTTIÖNGKMSSTIVSEDF-----ILPVY 58
DB 647 ISQITESTENMSLANK---EIQDLVDSKSDISIKLGEKSRILAEERFKILSTNLDLT 703
QY 59 KGE---LEKGYQDGEHISGFEGKKA-----GVINISKDTFIKPVFKIEK----- 104
DB 704 KEMDQLRRKFDV---LONTILKÖDSKTHETINEVSCSKSIYETELINKKEÖKUR 759
QY 105 -----KEENKPTFDVSKKDNPOVNSQLNESHK-KEDL---OREHSGQSDSTKDY 153
DB 760 VHEKRLKQELNK-----LSPKDSLRIMWTÖTÖLÖKEREDLLEFRKSCÖQKIDLEBDA 815

QY 154 TA-----TVLDKNNISSKSTNNPN 173
DB 816 LSELKKEITSÖKHITIKÖLEEDNNSN 840

RESULT 2
TOP2 PLAFK STANDARD; PRT; 1398 AA.
ID TOP2 PLAFK
AC P41001;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2.
OS Plasmodium falciparum (isolate K1 / Thailand).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCB1_TaxID=5839;
RN [1]
RX MEDLINE=94316496; PubMed=8041616;
RA Cheesman S., McAlleese S., Goman M., Johnson D., Horrocks P.,
RA Ridley R.G., Kilbey B.J.;
RT "The gene encoding topoisomerase II from Plasmodium falciparum.";
RL Nucleic Acids Res. 22:2547-2551(1994).
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -1- SUBUNIT: Homodimer (by similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -1- SIMILARITY: Belongs to the type II topoisomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X79345; ; NOT_ANNOTATED_CDS.
CC HSSP; P06786; 1BGW
CC InterPro; IPR003594; ATPbind_ATPase.
CC InterPro; IPR003957; CBFA_NFYB_topis.
CC InterPro; IPR001241; DNA_topoisomI.
CC InterPro; IPR002205; DNA_topoisomIV.
CC Pfam; PF00204; DNA_gyrasab; 1.
CC Pfam; PF00521; DNA_topoisomIV; 1.
CC Pfam; PF02518; HATPase_c; 1.
CC PRINTS; PR00615; CCAATSUBUNTA.
CC PRINTS; PR00418; TP12FAMILY.
CC ProDom; PD000742; DNA_topoisomIV; 1.

```

```

DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00433; TOP2C; 1.
DR SMART; SM00434; TOP4C; 1.
DR PROSITE; PS00177; TOPOISOMERASE II; 1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
FT NP_BIND 144 149 ATP (POTENTIAL).
FT ACT_SITE 830 830 DNA_CLEAVAGE (BY SIMILARITY).
FT DOMAIN 271 281 POLY-ASN.
FT DOMAIN 308 316 POLY-ASN.
FT DOMAIN 1089 1093 POLY-LYS.
FT DOMAIN 1227 1234 POLY-LYS.
SQ SEQUENCE 1398 AA; 161029 MW; BAADIBEE88FE5BE9 CRC64;

Query Match      11.0%; Score 99; DB 1; Length 1398;
Best Local Similarity 22.2%; Pred. No.12;
Matches 44; Conservative 45; Mismatches 67; Indels 42; Gaps 9;

QY 1 KIVKDPAR-----NTVKEPFLNKDVGSEELKPHRVTTIÖNGKMSSTIVSEEE 51
DB 1093 KVLVEELVYKGYDPYKDINKIKKEEIFEQELLDAD-NPE-----DNEELIAGITVXDY 1145
QY 52 DFIL--PVYKGLKGYQDGEHISGFEGKKNAGVINISKDTFIKPVFKIE----- 103
DB 1146 DYLLSMPIFSLTLEK---VEDLLTÖLKEKERLEILRNITVETMWLKDIEKVEALIEFOR 1202
QY 104 -----KEENKPTFDVSKKDNPOVNSQLNESHKEDLÖREHSGQ---KSDSTKDYTA 155
DB 1203 NVELNSREBSNK--FKVARKQ-----GRSMKKKKKKKKLSDSESGDSDSSEFLVN 1255
QY 156 TVLDKNNISSKSTNNPN 173
DB 1256 TLNKKNTKTKTTSSNN 1273

RESULT 3
YHO0 YEAST STANDARD; PRT; 1345 AA.
ID YHO0 YEAST
AC P38800;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 149.7 kDa protein in IRL-KSP1 intergenic region.
GN YHR080C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCB1_TaxID=4932;
RN [1]
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Faveille A., Fulton L., Gattung S., Geisel C., Kistren J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VII.1.";
RL Science 265:2077-2082(1994).
CC -1- SIMILARITY: TO YEAST YFL042C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U10556; AAB68895.1; -
CC PIR; S46817; S46817.

```

DR SGP: S0001122; YHR080C.  
 DR InterPro: IPR004182; GRAM\_dom.  
 DR Pfam: PF02893; GRAM\_1.  
 DR SMART: SM00568; GRAM\_1.  
 DR Hypothetical protein; Transmembrane.  
 FT TRANSMEM 1198 1218 POTENTIAL.  
 SQ SEQUENCE 1345 AA; 149679 MW; 2FDAB94A686564C2 CRC64;

Query Match 10.9%; Score 97.5; DB 1; Length 1345;  
 Best Local Similarity 27.0%; Pred. No. 15;  
 Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

QY 59 KGELEKQVQFQGWESGEGKK-DAGVYINLSKDTFVKPKIEKKEENKPTFEDSK 117  
 DB 1109 KGATKSG-----SVGGQKVSVDYIMSEKDI-----SRKSKKPKYKWK 1149

QY 118 KKDQNVNHSQLSHREKEDLQREHSQKSDSTKVATVLD--KNNISKSTN---NP 172  
 DB 1150 SHDKRPPHSKYE-----QKSESRKSDNDKQILHTILDVQNNFSSEIFNMKLLSP 1201

QY 173 NK 174  
 DB 1202 QK 1203

RESULT 4  
 MAPB\_HUMAN STANDARD; PRT; 2468 AA.  
 AC P46821;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP 1 light chain LC1].  
 GN MAP1B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RX MEDLINE=95104835; PubMed=7806212;  
 RA Lien L.V., Feener C., Fischbach N., Kunkel L.M.;  
 RT "Cloning of human microtubule-associated protein 1B and the  
 RT identification of a related gene on chromosome 15.";  
 RL Genomics 22:273-280(1994).  
 CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.  
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES  
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST  
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS  
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN  
 CC STABILIZING MICROTUBULES.  
 CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE  
 CC WITH MAP1A AND MAP1B PROTEINS.  
 CC -1- DOMAIN: Has a highly basic region with many copies of the sequence  
 CC KKEE and KKEI/V, repeated but not at fixed intervals, which is  
 CC responsible for the binding of MAP1B to microtubules.  
 CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated  
 CC from MAP1B by proteolytic processing. It is free to associate with  
 CC both MAP1A and MAP1B. It interacts with the amino-terminal region  
 CC of MAP1B (by similarity).  
 CC -1- SIMILARITY: TO MAP1A.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, L06237; AAA18904.1; -

DR Genew; HGNC:6836; MAP1B.  
 DR MIM: 157129; -  
 DR GO: GO:0005875; C:microtubule associated complex; TAS.  
 DR InterPro: IPR00102; MAP1B\_neuraxin.  
 DR Pfam: PF00414; MAP1B\_neuraxin; 10.  
 DR PROSITE: PS00230; MAP1B\_NEURAXIN; 6.  
 KW Microtubules; Repeat; Phosphorylation.  
 FT CHAIN ? 2468  
 FT REPEAT 1878 1894 MAP1B 1.  
 FT REPEAT 1895 1911 MAP1B 2.  
 FT REPEAT 1912 1928 MAP1B 3.  
 FT REPEAT 1929 1945 MAP1B 4.  
 FT REPEAT 1946 1962 MAP1B 5.  
 FT REPEAT 1963 1979 MAP1B 6.  
 FT REPEAT 1997 2013 MAP1B 7.  
 FT REPEAT 2014 2030 MAP1B 8.  
 FT REPEAT 2031 2047 MAP1B 9.  
 FT REPEAT 2048 2064 MAP1B 10.  
 FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY  
 KKEE AND KKEI/V REPEATS).  
 SQ SEQUENCE 2468 AA; 270618 MW; 540839C8D9D9d61 CRC64;

Query Match 10.9%; Score 97.5; DB 1; Length 2468;  
 Best Local Similarity 24.8%; Pred. No. 29;  
 Matches 36; Conservative 29; Mismatches 47; Indels 33; Gaps 6;

QY 14 KEPIANKDTGEVSELEKPRRTVTITONGKEMSTVSEDFLLPYKGELEKQVQFQGW 73  
 DB 584 EKVMWKDKDPVKTEKPSVTEKEVPSKEPS-----PV-KAVVA-----EK 623

QY 74 SGFEKGDAGVYINLSKDTFVKPKIEKKEENKPTFVSKKDKQNVNHSQLSH 133  
 DB 624 QATDVKPPRAAEKVKYKKEIKYK-----EDKKEKKEKKEVAKKEDTPT---KKEKP 675

QY 134 RKEDLQRE-----EHSQKSDSTKV 153  
 DB 676 KKEVYKKEVKKKEIKKEKKEPKKEV 700

RESULT 5  
 GYRA\_STABP STANDARD; PRT; 893 AA.  
 AC P54112;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA gyrase subunit A (EC 5.99.1.3).  
 GN GYRA OR SE0005.  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=1282;  
 RX MEDLINE=92102204; PubMed=1662027;  
 RA Sreedharan S., Peterson L.R., Fisher L.M.;  
 RT "Clprofloxacin resistance in coagulase-positive and -negative  
 RT staphylococci: role of mutations at serine 84 in the DNA gyrase A  
 RT protein of Staphylococcus aureus and Staphylococcus epidermidis.";  
 RL Antimicrob. Agents Chemother. 35:2151-2154(1991).  
 CC -1- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-  
 CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE  
 CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED  
 CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.  
 CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining  
 CC of double-stranded DNA.  
 CC -1- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA  
 CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE





DR InterPro: IPR004868; DNA pol B\_2.  
 DR Pfam: PF01175; DNA\_pol\_B\_2; 1.  
 DR SMART: SM00486; POLB; 1.  
 DR PROSITE: PS00116; DNA\_POLYMERASE\_B; 1.  
 DR Transferrase; DNA-directed DNA polymerase; DNA replication;  
 KW DNA-binding; Plasmid; Mitochondrion.  
 SQ SEQUENCE 1202 AA; 138279 MW; 51D41FCBDBF2CDE CRC64;

Query Match 10.6%; Score 95; DB 1; Length 1202;

Best Local Similarity 22.8%; Pred. No. 20;

Matches 44; Conservative 37; Mismatches 80; Indels 32; Gaps 9;

QY 10 NTTFVEFLNK---DTGSESELKPHRVTVTITONGKENSSTI---VSEEDF--ILPVYKG 60  
 DB 305 NTFQFFVYVNAKIKRPTGNVRSIGFQNT-TITDKETLIKTLAFLEREDITWVSDEG 363  
 QY 61 ELEK-----GYQPDGWEISGFEKKDAGYVNLKSDTFIKPKIE-----EKKEE 108  
 DB 364 DIDEKFPKGLSLSPFQPKLTKETGRVANYTPPIKKDIYVKQINKKINPGLDLPKTMDL 423  
 QY 109 NK-PFVDVSKKDNQVNHSQLNESHKEDLQREHSQKSDSTKQVATVLDKNI----- 163  
 DB 424 SKPPLKLNKQKTSSEIRMTIKKNQSYDI--IGHMINDENYITFRRAVDNSIIKIF 481  
 QY 164 ---SSKSTTNPN 173  
 DB 482 TVTDSMGNTNDPN 494

## RESULT 8

HS70\_PYRSA STANDARD; PRT; 649 AA.

AC P37899;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Heat shock 70 kDa protein.  
 GN HSP70.  
 OS Pyrenomonas salina.  
 OG Nucleomorph.  
 OC Eukaryota; Cryptophyta; Cryptomonadaceae; Pyrenomonas.  
 OX NCBI\_TaxID=3034;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94268506; Pubmed=8208251;  
 RA Hofmann C.J.B., Rensing S.A., Haebler M.M., Martin W.F., Mueller S.B.,  
 RA Couch J., McFadden G.I., Igloi G.L., Maier U.-G.;  
 RT "The smallest known eukaryotic genomes encode a protein gene: towards  
 RT an understanding of nucleomorph functions."  
 RL Mol. Gen. Genet. 243:600-604(1994).  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X72621; CA51197.1; -  
 DR PIR: S42488; S42488.  
 DR HSSP: P08109; ICR.  
 DR InterPro: IPR001023; Hsp70.  
 DR Pfam: PF00012; HSP70; 1.  
 DR PRINTS: PR00301; HEATSHOCK70.  
 DR Prodom: PD000089; HSP70; 1.  
 DR PROSITE: PS00297; HSP70\_1; 1.  
 DR PROSITE: PS00329; HSP70\_2; 1.  
 DR PROSITE: PS01036; HSP70\_3; 1.  
 KW ATP-binding; Heat shock; Nucleomorph.  
 KW SEQUENCE 649 AA; 72079 MW; B627B08FP90C9164 CRC64;

Query Match 10.4%; Score 93.5; DB 1; Length 649;  
 Best Local Similarity 24.5%; Pred. No. 13;  
 Matches 38; Conservative 26; Mismatches 50; Indels 41; Gaps 7;

QY 17 ILNKDTGSESELKPHRVTVTITONG--XEMSTIVSEDFILPVYKGELEKGYQPDGWEI 73  
 DB 491 ILNVASDPSKTSKSKNKTITNDKGLSKKEIERMVEAE-----KYKTEDEK----- 537  
 QY 74 SGFEKKDA-----GYVNLN---KDTFIKVPFKKIEKEEENKPTVDVSKKDNQVNH 125  
 DB 538 -LDKKELAKNSLENVAVNINRYVD-----EKLKKEIOEDKKSIEKYE----- 582  
 QY 126 HSQLNESHKEDLQREHSQKSDSTKQVATVLDK 160  
 DB 583 --VLEFITNEDLEKEEYEEKEELKANPANIPIK 615

## RESULT 9

ADSV\_HUMAN STANDARD; PRT; 715 AA.

AC Q9YU03; Q9YU07;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Adseverin (Scinderin).  
 GN SCIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=22388257; Pubmed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heisch F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrinai P., Prange C.,  
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Roark S.A., McKernan P.J., McKernan K.T., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Heltan E., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE OF 1-527 FROM N.A.  
 RA Kallicki J., Smith-Craig R.;  
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 248-715 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Isogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,  
 RA Arima M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
 RA Wakamatsu A., Nakamura Y., Nagahari K., Masuo Y., Oshima A.;  
 RT "NERO human cDNA sequencing project";  
 RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: Ca(2+)-dependent actin filament-severing protein that is  
 CC presumed to have a regulatory function in exocytosis by affecting  
 CC the organization of the microfilament network underneath the  
 CC plasma membrane. In vitro, also has barbed end capping and  
 CC nucleating activities in the presence of Ca(2+).

```

CC -!- SIMILARITY: BELONGS TO THE VILILIN/GELSOLIN FAMILY.
CC -!- SIMILARITY: Contains 6 gelsolin-like repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; BC021090; AAH21090.1; -
CC EMBL; AC005281; AAD15423.1; -
CC EMBL; AK027778; BAB55361.1; -
CC HSP; P02640; 2VIL;
CC InterPro; IPR001974; Gelsolin.
CC Pfam; PF00626; Gelsolin; 6.
CC PRINTS; PR00597; GELSOLIN.
CC SMART; SM00262; GEL; 6.
CC Cytochrome; Actin-binding; Repeat; Calcium; Gapping protein.
CC ACTIN-SEVERING (POTENTIAL).
CC CA(2+)-DEPENDENT ACTIN BINDING.
CC DOMAIN 1 363
CC FT REPEAT 364 715 GELSOLIN-LIKE 1.
CC FT REPEAT 27 76 GELSOLIN-LIKE 1.
CC FT REPEAT 148 188 GELSOLIN-LIKE 2.
CC FT REPEAT 265 307 GELSOLIN-LIKE 3.
CC FT REPEAT 323 398 GELSOLIN-LIKE 4.
CC FT REPEAT 523 564 GELSOLIN-LIKE 5.
CC FT REPEAT 626 668 GELSOLIN-LIKE 6.
CC FT SITE 112 119 POLYPHOSPHONOSITIDE BINDING (BY
CC FT SITE 138 146 POLYPHOSPHONOSITIDE BINDING (BY
CC FT SITE 146 146 POLYPHOSPHONOSITIDE BINDING (BY
CC FT CONFLICT 61 61 R -> H (IN REF. 2).
CC FT SEQUENCE 715 AA; 80508 MW; 45FBE42CBCEFDDB0 CRC64;
SQ
Query Match 10.4%; Score 93; DB 1; Length 715;
Best Local Similarity 21.0%; Pred. No. 16;
Matches 37; Conservative 39; Mismatches 58; Indels 42; Gaps 6;
QY 3 VKKDFARNTTVEPIINIKDTGEVSEIKPHRVYTTTONGKEMSGTIVSEEDFILPYVKGEL 62
DB 239 ITADISNRKAKTYWSDASGSM-----RTVVAENEPFMMAMLLSECCFLID--HGAA 290
QY 63 EKGYPDGEIGSEIFSEK-----KDAGYVINSKDFEIK-----PVFKK----- 100
DB 291 KQIFWKGKQANPQERKAAKTAEBFLQOMNTYSKNTQIVLPBGGETPIFKQPFKQWRDK 350
QY 101 -----IEKKEENKPTFDVSKKKDNQV--NHSQLESNHRKEDLOREHS 144
DB 351 DSDGKGYVTEKVAQIKQIPFDASKLSSPQMAQHNMVDDGSGKVAIMVENN 406
RESULT 10
MAPB MOUSE STANDARD; PRT; 2464 AA.
AC P14873;
DT 01-APR-1990 (rel. 14, Created)
DT 01-APR-1990 (rel. 14, Last sequence update)
DT 28-FEB-2003 (rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))
DE [Contains: MAP1 light chain Ict1].
DE GN MAP1B OR MAP1B OR MAP1B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A., AND DOMAIN.
RC STRAIN=Swiss Webster; TISSUE=Brain;
RA MEDLINE=90094539; PubMed=2480963;
RX Noble M., Lewis S.A., Cowan N.J.;
RT "The microtubule binding domain of microtubule-associated protein
RT MAP1B contains a repeated sequence motif unrelated to that of MAP2

```

```

RT and tau."
RL J. Cell Biol. 109:3367-3376(1989).
CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -!- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC OF MAP1B.
CC -!- SIMILARITY: TO MAP1A.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X51386; CAA35761.1; -
CC PIR; S07549; QRMSP1.
CC MGD; MG1:1306778; Metap1b.
CC DR GO:0016358; P.dendrite morphogenesis; IMP.
CC DR GO:0001578; P.microtubule bundling; IMP.
CC DR InterPro; IPR00102; MAP1B neuraxin.
CC DR Pfam; PF00414; MAP1B neuraxin; 10.
CC DR PROSITE; PS00230; MAP1B_NBDAXIN; 7.
CC KW Microtubules; Repeat; Phosphorylation.
CC FT CHAIN ? 2464
CC FT REPEAT 1874 1890 MAP1 1.
CC FT REPEAT 1891 1907 MAP1 2.
CC FT REPEAT 1908 1924 MAP1 3.
CC FT REPEAT 1925 1941 MAP1 4.
CC FT REPEAT 1942 1958 MAP1 5.
CC FT REPEAT 1959 1975 MAP1 6.
CC FT REPEAT 1993 2009 MAP1 7.
CC FT REPEAT 2010 2026 MAP1 8.
CC FT REPEAT 2027 2043 MAP1 9.
CC FT REPEAT 2044 2060 MAP1 10.
CC FT DOMAIN 589 787
CC FT SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFBDBA87 CRC64;
SQ
Query Match 10.4%; Score 93; DB 1; Length 2464;
Best Local Similarity 27.8%; Pred. No. 61;
Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;
QY 14 KEPIINIKDTGEVSEIKPHRVYTTTONGKEMSGTIVSEEDFILPYVKGELKGYQFDGWEI 73
DB 584 EKVLYKKDKPVKTESKP--SVT---EKVSS---KEE--SPV-KAEVA-----EK 623
QY 74 SGFEGKQAGVYINISKDTFIKPVK-KIEKKEENKPTFDVSKKKDNQVNHNSQLES 132
DB 624 QATESKP-----KTKDKVVKKEIKTKLEKKEE--RPKEVVKKEDKTPL---KDEK 672
QY 133 HKKEDLORE-----EHSQSDSKVDTATYATLDGNISSTKTTNNPK 174
DB 673 PRKEVYKKEIKKEIKKEERKELKEVKKETPLKDAKKEVKEEKEVKEEKEPKK 728
RESULT 11
DACA BACSU STANDARD; PRT; 443 AA.
ID DACA BACSU
AC P08750;
DT 01-AUG-1988 (rel. 08, Created)

```

DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE D-alanyl-D-alanine carboxypeptidase precursor (EC 3.4.16.4) (DD-  
 DE peptidase) (DD-carboxypeptidase) (CPase) (PAPS).  
 GN DACC.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 CX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=96051385; PubMed=7584024;  
 RA Ogasawara N., Nakai S., Yoshikawa H.;  
 RT "Systematic sequencing of the 180 kilobase region of the Bacillus  
 RT subtilis chromosome containing the replication origin.";  
 RL DNA Res. 1:1-14(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=96044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Berteiro M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Dentzer F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabre C., Ferrati E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,  
 RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kashara Y., Klaerr-Blanchard N., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kuitica K., Lapidis A., Lardinois S., Lamber J., Lazarovic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Priesean E., Puig P., Fumelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Setiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarelli A.,  
 RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yaman K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumslein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis.";  
 RL Nature 390:249-256(1997).  
 RN [3]  
 RP SEQUENCE OF 32-102.  
 RX MEDLINE=80182289; PubMed=6768745;  
 RA Waxman D.J., Strominger J.L.;  
 RT "Sequence of active site peptides from the penicillin-sensitive D-  
 RT alanine carboxypeptidase of Bacillus subtilis. Mechanism of  
 RT penicillin action and sequence homology to beta-lactamases.";  
 RL J. Biol. Chem. 255:3964-3976(1980).  
 RN [4]  
 RP SEQUENCE OF 103-443 FROM N.A.  
 RX MEDLINE=86250602; PubMed=3087956;  
 RA Todd J.A., Roberts A.N., Johnstone K., Piggett P.J., Winter G.,  
 RA Ellar D.J.;  
 RT "Reduced heat resistance of mutant spores after cloning and  
 RT mutagenesis of the Bacillus subtilis gene encoding penicillin-binding  
 RT protein 5.";  
 RL J. Bacteriol. 167:257-264(1986).  
 RN [5]  
 RP SEQUENCE OF 414-443.  
 RX MEDLINE=81117303; PubMed=6780559;  
 RA Waxman D.J., Strominger J.L.;  
 RT "Primary structure of the COOH-terminal membranous segment of a

RT penicillin-sensitive enzyme purified from two Bacilli.";  
 RL J. Biol. Chem. 256:2067-2077(1981).  
 CC - FUNCTION: REMOVES C-TERMINAL D-ALANYL RESIDUES FROM SUGAR-PEPTIDE  
 CC CELL WALL PRECURSORS.  
 CC - CATALYTIC ACTIVITY: D-alanyl-D-alanine + H(2)O = 2 D-alanine.  
 CC - PATHWAY: Peptidoglycan synthesis; final stages.  
 CC - SUBCELLULAR LOCATION: Membrane-associated.  
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S11; ALSO KNOWN AS THE  
 CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 1 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: D26185; BAA05246.1; -;  
 CC DR EMBL: Z99104; CAB11786.1; -;  
 CC DR EMBL: M13766; AAA22375.1; -;  
 CC DR PIR: S66040; S66040.  
 CC DR MEROPS: S11.001; -;  
 CC DR Subtilist; BG10074; dacc.  
 CC DR InterPro: IPR001967; Ala/AlaCpIase1.  
 CC DR Pfam: PF00768; Peptidase S11; 1.  
 CC DR PRINTS: PR00725; DA0ACBPTASE1.  
 CC DR Hydrolase; Carboxypeptidase; Peptidoglycan synthesis; Cell wall;  
 CC Membrane; Signal; Complete proteome.  
 CC KW SIGNAL 1 31  
 CC FT CHAIN 32 443 D-ALANYL-D-ALANINE CARBOXYPEPTIDASE.  
 CC FT ACT SITE 67 67 ACYLATED BY PENICILLIN.  
 CC FT CONFLICT 100 100 E -> Q (IN REF. 3).  
 CC FT CONFLICT 227 227 E -> Q (IN REF. 4).  
 CC SQ SEQUENCE 443 AA; 48635 MW; DA6C5B0307D7C117 CRC64;  
 CC -----  
 CC Query Match 10.3% Score 92.5; DB 1; Length 443;  
 CC Best Local Similarity 25.0%; Pred. No. 10; Indels 33; Gaps 6;  
 CC Matches 36; Conservative 26; Mismatches 49;  
 CC QY 6 DPA-RNTVKEPFLNKDTGSEVSELPKPRVTVTTONGKMSSTIVSEBDFILPVYKGELEK 64  
 CC DB 304 DYAPNPFMKRIYABGD-----QYKGRK-TISVDKGEKEKGIYTNKAFSLPVKNGE-EK 356  
 CC QY 65 GYQPDGWEISGPEEGKQAGVYINLSKDTFIPVFKLIEKKBEENKPTFDVSKKQNDPOV 124  
 CC DB 357 NYKAK-----VTINKDMLTAPVKKGTGVK-----LTAEYTGDEK 391  
 CC QY 125 NNSQUNESHKEDLQREHSQKSD 148  
 CC DB 392 DYGLNSDLADVLVTKEVSKAN 415  
 CC -----  
 CC RESULT 12  
 CC IGA2\_HAEMIN STANDARD; PRT; 1702 AA.  
 CC AC P45384;  
 CC DT 01-NOV-1995 (Rel. 32, Created)  
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 CC DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).  
 CC GN IGA.  
 CC OS Haemophilus influenzae.  
 CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 CC OC Pasteurellaceae; Haemophilus.  
 CC CX NCBI\_TaxID=727;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=HK715 / Serotype B;  
 CC RX MEDLINE=92234949; PubMed=1373717;  
 CC RA Poulsen K., Reinholdt J., Kilian M.;  
 CC RT "A comparative genetic study of serologically distinct Haemophilus  
 CC influenzae type 1 immunoglobulin A1 proteases.";

```

RL J. Bacteriol. 174:2913-2921 (1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTERACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; M87489; AAA24966.1; -
CC PIR; A41859; A41859.
CC MEROPS; S06.001; -.
CC InterPro: IPR006315; Autotransport.
CC InterPro: IPR005546; Autotransporter.
CC InterPro: IPR000710; IGA_S6.
CC InterPro: IPR004899; Peractin.
CC Pfam; PF03797; Autotransporter; 1.
CC Pfam; PF02395; IGA1; 1.
CC Pfam; PF03212; Peractin; 1.
CC PRINTS; PR00921; IGASERPRASE.
CC TIGRfam; TIGR01414; autotrans_bar1; 1.
CC HydroLase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.
CC SIGNAL 1 25 IMMUNOGLOBULIN A1 PROTEASE.
CC CHAIN 26 1014 HELPER PEPTIDE (POTENTIAL).
CC PROPE 1015 1702 PROBABLE
CC ACT SITE 288 288
CC DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
CC FT REPEAT 1109 1116 .1.
CC FT REPEAT 1117 1124 2.
CC FT REPEAT 1117 1124 2.
CC SQ SEQUENCE 1702 AA; 186539 MW; 860F70D267807A6 CRC64;

Query Match 10.3%; Score 92.5; DB 1; Length 1702;
Best Local Similarity 23.9%; Pred. No. 44;
Matches 44; Conservative 26; Mismatches 101; Indels 13; Gaps 5;

QY 2 IYVKDFARNT---TYKEFLINKDGEVSEELKPHRVTVTIQNGKEMSSTIVSE---EDPI 54
DB 1206 VVSKNTEHTTDOPTREKTAKVETKTQE--PPVVASQSPQEOSEFVQPAVLESN 1263
QY 55 LPVYKGELEKGYQFDGMEISGEGKDA-GYVINLSKDTPIKPVFKIEKEENKPTF 113
DB 1264 VPTVNAEVEVQQLQVTSATVSTKQPADENSINTSATAITETAKSKQPOETASTRE 1323
QY 114 DYSKKKDNQVNHSQLNESHREKDLQREHS--QKSDTKDYATVADKNNISSKSTN 170
DB 1324 DASQHKANTVADNSVANNSESSSEPKRRRRRISIQPOETSAEETTAASTDETTIADNSKS 1383
QY 171 NPNK 174
DB 1384 KPR 1387

RESULT 13
YDT2_SCHPO STANDARD; PRT; 1888 AA.
ID_YDT2_SCHPO
AC 014207;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)

```

```

DE Hypothetical protein C6B12.02c in chromosome I.
GN SPAC6B12.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972.
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean U.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Gymnopoulos B.,
RA Wellfens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Fuchs M., Fritze C., Holzer E., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wandt R., Purnelle B.,
RA Goftau A., Cadieu E., Dreano S., Gloux S., Leleure V., Moutier S.,
RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,
RA "The genome sequence of Schizosaccharomyces pombe."
RA Nature 415:871-880(2002).
RT
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; Z98531; CAB11064.1; -
CC PIR; T39009; T39009.
CC DR GeneDB SPombe; SPAC6B12.02c; -.
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 657 717 POTENTIAL.
CC FT TRANSMEM 755 775 POTENTIAL.
CC FT TRANSMEM 866 886 POTENTIAL.
CC FT TRANSMEM 915 935 POTENTIAL.
CC FT TRANSMEM 977 997 POTENTIAL.
CC FT TRANSMEM 1033 1053 POTENTIAL.
CC FT TRANSMEM 1336 1356 POTENTIAL.
CC FT TRANSMEM 1645 1665 POTENTIAL.
CC FT DOMAIN 1662 1665 POLY-LEU.
CC SQ SEQUENCE 1888 AA; 217432 MW; 8AD3BBCE32397C29 CRC64;

Query Match 10.3%; Score 92.5; DB 1; Length 1888;
Best Local Similarity 23.1%; Pred. No. 49;
Matches 39; Conservative 25; Mismatches 76; Indels 29; Gaps 5;

QY 11 TTYKEFLINKDGEVSEELKPHRVTVTIQNGKEMSSTIVSE---EDPI 69
DB 394 TSENFPQNVANNAVSTIPVATTTTKMKQRKFVVEVKEKLPDLLEST----- 442
QY 70 GWEISGEGKDAQGVINLSKDTPIKPVFKIEKEENKPTFVSKKKNQVNHSQL 129
DB 443 -----GKAPKFLRVAFARSSSHIP--KMIIRKQMDSKKYFSPKESDROVIDOVL 492

```

QY 130 NESHKEDLOREHSOKSDS-TKDYATVLDKN-----NISKSTTN 171  
 DB 493 DWYSGKHELVOGSHSYKPKSDSKVGNIFSVNSKSHSNINAKTAAVN 541

RESULT 14

NSBL\_MOUSE STANDARD; PRT; 406 AA.  
 AC Q9ULJ5; 088832; Q8VC71; Q9CUM1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Nucleosome binding protein 1 (Nucleosome binding protein 45) (NBP-45) (GARP45 protein).  
 GN NSBP1 OR GARP45.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;

[1]  
 SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
 RX MEDLINE=20158948; PubMed=10692437;  
 RA Shitakawa H., Landman D., Postnikov V.V., Bustin M.;  
 RT "NBP-45, a novel nucleosomal binding protein with a tissue-specific and developmentally regulated expression.";  
 RL J. Biol. Chem. 275:6368-6374(2000).

[2]  
 SEQUENCE FROM N.A.  
 RA Onoda G., Suzuki N., Saito H., Honda T., Sato H., Kuwano R.;  
 RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.

[3]  
 SEQUENCE FROM N.A.  
 RC TISSUE=Colon;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Ueda T.B., Toshiyuki S., Carrinci P., Prange C., Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hilyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalski U., Smallus D.E., Scherch A., Schin J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[4]  
 SEQUENCE OF 7-195 FROM N.A.  
 RP STRAIN=C57BL/6J; TISSUE=Hippocampus;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arai K., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamana I., Saito T., Okazaki Y., Gojohori T., Bono H., Kaubawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H., Kuehl P., Lewis S., Maeno Y., Nikaide I., Pesole G., Quackenbush J., Schriml L.M., Steubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldairelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M., Guttingich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Nombarts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kontani S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).

CC -1- FUNCTION: Binds specifically to nuclear core particles and acts as  
 a transcriptional activator.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Widely expressed with highest levels in  
 submaxillary gland, thymus, kidney and liver and lowest levels in  
 brain, lung, pancreas and eye.  
 CC -1- DEVELOPMENTAL STAGE: Highest levels are found in 7-day-old  
 embryos. Levels in the 7-day-old embryo are 4-fold higher than in  
 the adult and almost 10-fold higher than in later embryonic  
 stages.

CC -1- SIMILARITY: BELONGS TO THE HMG-14/HMG-17 PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC EMBL; AF213454; AAF30179.1; -  
 CC EMBL; AB018374; BAA33783.2; -  
 CC EMBL; BC021626; AAB21626.1; -  
 CC EMBL; AK013748; BAB28982.1; -  
 CC MGD; MGI:1355295; Nsdp1.  
 CC GO; GO:0005654; C:nucleoplasm; IDA.  
 CC GO; GO:0003682; F:chromatin binding activity; IDA.  
 CC GO; GO:0006356; P:regulation of transcription from Pol I prom. .; IDA.  
 CC InterPro; IPR000079; HMG\_14\_17.  
 CC Pfam; PF01101; HMG14\_17; 1.  
 CC SMART; SM00527; HMG17; 1.  
 CC PROSITE; PS00355; HMG14\_17; FALSE NEG.  
 CC Transcription regulation; Activator; DNA-binding; Nuclear protein.  
 CC CONFLICT 74 74 M -> V (IN REF. 1).  
 CC CONFLICT 390 390 N -> H (IN REF. 2).  
 CC SEQUENCE 406 AA; 45344 MW; 59A4305613EC9679 CMC64;

Query Match 10.3%; Score 92; DB 1; Length 406;  
 Best Local Similarity 27.4%; Pred. No. 10;  
 Matches 48; Conservative 20; Mismatches 59; Indels 48; Gaps 9;

QY 1 KIIVKDPKANTVVEFLINDTGEVSELRPRVVTIONK-----EMSTVSEEDF 53  
 DB 126 KANTDVEKDDGE--HKDTGE-----EVDGKIEEGKNGKGTAKSD-- 169

QY 54 ILPVYKLEKGYPDGEISGFEKKDAGVINLSKD-----TFIKPVFKLIEK 104  
 DB 170 -AEVSKDEEKGDKG-----EDGKEGDEKDEKDGKDTGTEVVEGQKKEAED 222

QY 105 ---KEENKPTFDVSKKKNPQVNHSQLNESHKEDLQREE--HSQKSDSTKDV 153  
 DB 223 DGKCKEENK---EVGKEGQPEEDGKEDLHREVGKEDLHEDGKGEQPEEDGKEI 274

RESULT 15

SPT7 YEAST STANDARD; PRT; 1332 AA.  
 AC P35177;  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Transcriptional activator SPT7.  
 GN SPT7 OR YBR081C OR YBR0739.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

```

OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=8288C;
RC MEDLINE=95229044; PubMed=7713415;
RA Ganeheroff L.J., Dollard C., Tan P., Winston F.;
RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein
RL important for transcription in vivo.";
RL Genetics 139:523-536(1995).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=8288C;
RC MEDLINE=95076715; PubMed=7985423;
RA van der Aart O.J.M., Barthe C., Doignon F., Aigle M., Crouzet M.,
RA Steensma H.Y.;
RT "Sequence analysis of a 31 kb DNA fragment from the right arm of
RL Saccharomyces cerevisiae chromosome II.";
RL Yeast 10:959-964(1994).
RN [3]
RN SEQUENCE OF 1-835 FROM N.A.
RC STRAIN=8288C;
RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestearazu A.,
RA Vissers S.;
RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RN SEQUENCE OF 463-523 FROM N.A.
RC MEDLINE=92285152; PubMed=1350857;
RA Haynes S.R., Dollard C., Winston F., Beck S., Trowdale J.,
RA David I.B.;
RT "The bromodomain: a conserved sequence found in human, Drosophila and
RL yeast proteins.";
RL Nucleic Acids Res. 20:2603-2603(1992).
CC - FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY
CC OTHER GENES.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - SIMILARITY: Contains 1 bromodomain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, L22537; AAC37424.1; -
DR EMBL, X76294; CAAS3940.1; -
DR EMBL, Z35950; CAAS5026.1; -
DR EMBL, M87651; AAA35087.1; -
DR PIR, S41552; S41552.
DR HSSP, Q92831; 1B91.
DR TRANSFAC, T04835; -
DR SGD, S0000285; SPT7.
DR GO, GO:0000124; C:SAGA complex; IDA.
DR InterPro, IPR001487; Bromodomain.
DR Pfam, PF00439; bromodomain; 1.
DR PRINTS, PRO0297; BROMODOMAIN.
DR SMART, SM00297; BROMO.1.
DR PROSITE, PS00653; BROMODOMAIN_1; 1.
DR PROSITE, PS50014; BROMODOMAIN_2; 1.
DR Transcription regulation; Nuclear protein; Activator; Bromodomain.
DR DOMAIN 458 528 BROMODOMAIN.
DR SEQUENCE 1332 AA; 152616 MW; 083B63624669244F CRC64;
SO

```

```

DB 538 KSLQIRIMPTITIRNADLEKEI---EDM-----EKDKYEDDEEEVAGSRKG-- 585
QY 84 YVINLSKDTFK--PVFKTIEEKEENKPFED-----VSKKD-----NP 122
DB 586 --LNNQAHMLAENGKSVKSSKTVKDBAFNDKLTSLVPEGEKDKTASSTVYVHE 643
QY 123 QVNHQSLNESHRKEDLOR-EESHQKSDSTKVATVLD-KNNISKSTTN 170
DB 644 NVNKNIEKNGKNEEQDWEESKTEDSSKADAKKOTEDLOKXTAEN 693

```

RESULT 16

ID	LIP STAMP	STANDARD;	PRT;	688 AA.
AC	002510;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Lipase precursor (EC 3.1.1.3) (Glycerol ester hydrolase).			
GN	GEHC OR SE0281.			
OS	Staphylococcus epidermidis.			
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.			
OX	NCBI_TaxID=1282;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 303-315.			
RC	STRAIN=9;			
RX	MEDLINE=93171870; PubMed=8436947;			
RA	Farrrell A.M., Foster T.O., Holland K.T.;			
RT	"Molecular analysis and expression of the lipase of Staphylococcus			
RL	epidermidis.";			
RL	J. Gen. Microbiol. 139:267-277(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 12228;			
RA	Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,			
RT	Chen Z., Wen Y.;			
RL	Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.			
CC	- CATALYTIC ACTIVITY: Triacylglycerol + H(2)O = diacylglycerol + a			
CC	fatty acid anion.			
CC	- SUBCELLULAR LOCATION: Secreted.			
CC	- MISCELLANEOUS: THE EXPRESSION OF STAPHYLOCOCCUS LIPASE IS			
CC	NEGATIVELY REGULATED BY BACTERIOPHAGE LYSOGENIZATION (LIPASE			
CC	CONVERSION).			
CC	- SIMILARITY: BELONGS TO THE ABHYDROLASE SUPERFAMILY. LIPASE FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL, M95577; AAA19729.1; -			
DR	EMBL, AE016744; AA003878.1; -			
DR	PIR, A47705; A47705.			
DR	InterPro, IPR005877; Gpos_Y5IRK.			
DR	InterPro, IPR000734; Lipase.			
DR	InterPro, IPR000379; Ser_estr_site.			
DR	Pfam, PF04650; Y5IRK_signal; 1.			
DR	TIGRFAME, TIGR01168; Y5IRK_signal; 1.			
DR	PROSITE, PS00120; LIPASE_SER; 1.			
KW	Hydrolase; Lipid degradation; Zymogen; Signal; Complete proteome.			
FT	SIGNAL 1 35			
FT	POTENTIAL.			
FT	PROPEP 36 302			
FT	REMOVED IN THE MATURE FORM.			
FT	CHAIN 303 688			
FT	LIPASE.			
FT	ACT_SITE 418 418			
FT	CHARGE RELAY SYSTEM (BY SIMILARITY).			
FT	ACT_SITE 648 648			
FT	CHARGE RELAY SYSTEM (BY SIMILARITY).			
FT	CONFLICT 96 96			
FT	W -> G (IN REF. 1).			
FT	E -> G (IN REF. 1).			
FT	CONFLICT 120 120			
SO	SEQUENCE 688 AA; 77343 MW; 6C95DBA7BAF86F6 CRC64;			



Query Match 10.0%; Score 90; DB 1; Length 688;  
 Best Local Similarity 24.4%; Pred. No. 25;  
 Matches 39; Conservative 30; Mismatches 65; Indels 26; Gaps 7;

QY 19 NKDTEVSELKPHRYVTIIONGKEMSTIVSEDFILPYKGLKGYOPDGMETISGFE- 77  
 DB 68 NKNVNEKSVN-----SITENESLHNERPKXEDWT-----QQQDSQNDKKSSEVWQ 115

QY 78 GKKGAGYVNLKDTFKPKVFKKIE-EKKEENKPTFDVSKKKDNPQV--NHSQLESNR 134  
 DB 116 NKENAFVQNHSEB---KPOQOEVELEKASENNQTLHSAQASNEDEVTKPESQDLNTAA 172

QY 135 KEDLQREHSOKSDSTKDTATVLDKNNTSSKSTTNPNK 174  
 DB 173 KOEDSOKENLTKDPTQSSKTTDL-----RATANOSK 205

RESULT 17  
 YFK8 YEAST STANDARD; PRT; 778 AA.

AC P43610;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical 88.7 kDa helicase in CDC26-SAP155 intergenic region.  
 GN YFR038W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=95400292; PubMed=7670463;  
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,  
 RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,  
 RA Yamazaki M., Tashiro H., Eki T.;  
 RT "Analysis of the nucleotide sequence of chromosome VI from  
 RT Saccharomyces cerevisiae.";  
 RL Nat. Genet. 10:261-268 (1995).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=96287654; PubMed=8666381;  
 RA Eki T., Naitou M., Hagiwara H., Abe M., Ozawa M., Sasanuma S.-I.,  
 RA Sasanuma M., Tsuchiya Y., Shibata T., Watanabe K., Ono A.,  
 RA Yamazaki M.-A., Tashiro H., Hanaoka F., Murakami Y.;  
 RT "Fifteen open reading frames in a 30.8 kb region of the right arm of  
 RT chromosome VI from Saccharomyces cerevisiae.";  
 RL Yeast 12:117-190 (1996).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, D50617, BAA09277.1; -  
 CC PIR, S56293, S56293.  
 CC SGD, S0001934, YFR038W.  
 CC InterPro: IPR001410; DEAD.  
 CC InterPro: IPR001650; Helicase\_C.  
 CC InterPro: IPR000330; SNF2\_N.  
 CC Pfam, PF00271, helicase\_C\_1.  
 CC Pfam, PF00176, snf2\_N\_1.  
 CC SMART, SM00487, DEXDC\_1.  
 CC SMART, SM00490, HELIC\_1.  
 KW Hypothetical protein; Nuclear protein; DNA-binding; Helicase;  
 KW ATP-binding.

FT NP BIND 247 254 ATP (POTENTIAL).  
 FT SITE 352 355 DEGH BOX.  
 SQ SEQUENCE 778 AA; 88730 MW; 3B6C0857B5EABD84 CRC64;

Query Match 10.0%; Score 90; DB 1; Length 778;  
 Best Local Similarity 23.0%; Pred. No. 28;  
 Matches 44; Conservative 29; Mismatches 76; Indels 42; Gaps 6;

QY 5 KDFANNTVKEPILNKDTGEVSELKPHRYVTIIONGKEMSTI-----IYSEDFILP 56  
 DB 26 RNMSSGVVYREKEVNDLTADISDSDDSDSEDNKHKGDNDTPATLQDVHSDBDI--- 82

QY 57 VKKGLKGYOPDGMETISGFEKKGAGYVNLKDTFKPKVFKKIEKKEENKPTFDVSK 116  
 DB 83 -----QDSDDSDTEAVQ-AQYVDKLAKDT--KSEOKSLDDELSEMDTIVSLK 129

QY 117 KKKDNPQVNHSQLNES-----HKEDLQK-----EHSOKSDSTKDTATATVL 158  
 DB 130 LKKLNEFVRQSGVYSIIADTLHRSNREYVANTKDNNSDDEHSSKKRKTKKSTIDF 189

QY 159 DKNNTSSKSTT 169  
 DB 190 FKKQKKNEDTT 200

RESULT 18  
 TIG STAM STANDARD; PRT; 433 AA.

AC Q99T16;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trigger factor (TF).  
 GN TIG OR SAV1675 OR SA1499 OR MM1619.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699),  
 OS Staphylococcus aureus (strain N315), and  
 OS Staphylococcus aureus (strain MM2).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxId=158878, 158879, 196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MU50 / ATCC 700699, and N315;  
 RX MEDLINE=2111952; PubMed=11418146;  
 RA Kuroda M., Ono T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekimizu K., Hirakawa H., Kohara S., Goto S., Yabuzaki J.,  
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayaishi H., Hiramatsu K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 RT aureus.";  
 RL Lancet 357:1225-1240 (2001).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MM2;  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwana N., Asano K., Naito T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiramatsu K.;  
 RT "Genome and virulence determinants of high virulence community-  
 RT acquired MRSA.";  
 RL Lancet 359:1819-1827 (2002).  
 CC -1- FUNCTION: Involved in protein export. Acts as a chaperone by  
 CC maintaining the newly synthesized protein in an open conformation  
 CC (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE FKBP-TYPE PRPLASE FAMILY. TIG SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF003363; BAB57837.1; -

CC EMBL; AF003363; BAB57837.1; -

DR EMBL; AF004827; BAB95484.1; -

CC EMBL; AF004827; BAB95484.1; -

DR PIR; A89951; A89951.

CC PIR; A89951; A89951.

DR HAMAP; MF\_00303; -; 1.

CC HAMAP; MF\_00303; -; 1.

DR InterPro; IPR001179; FKBP\_PPIase.

CC InterPro; IPR005215; Trl9\_fac.

DR Pfam; PF00254; FKBP; 1.

CC Pfam; PF00254; FKBP; 1.

DR TIGRfam; TIGR00115; tlg; 1.

CC TIGRfam; TIGR00115; tlg; 1.

DR PROSITE; PS00453; FKBP\_PPIase\_1; FALSE\_NEG.

CC PROSITE; PS00453; FKBP\_PPIase\_1; FALSE\_NEG.

DR PROSITE; PS00454; FKBP\_PPIase\_2; FALSE\_NEG.

CC PROSITE; PS00454; FKBP\_PPIase\_2; FALSE\_NEG.

KW Cell division; Chaperone; Isomerase; Rotamase; Complete proteome.

FT DOMAIN 163 248 PPIase, FKBP-TYPE.

ST SEQUENCE 433 AA; 8865D9AF6A1BC1E7 CRC64;

QY Query Match 10.0%; Score 89.5; DB 1; Length 433; Best Local Similarity 22.1%; Pred. No. 16; Mismatches 60; Indels 83; Gaps 13; Matches 50; Conservative 33;

DB 21 DTGEVSELKPHRYVT-IQNGKE--MSSTVSEEDFILPVYKG-ELEK----- 64

DB 81 DETDIPVAVQPEVSVTIQIEKQDFEATVTPREVKLDGYGLIEKQETELSDDELQ 140

QY 65 -----GYPDGMEISG--FEKKDAGYINISKDTFIRP 96

DB 141 AIDHSIGHLAEVWKEDGVENGDTVINIDFSG--SVDEGEFEGQAGCYLDLEIGSGFI-P 198

QY 97 VPK-----KIEEKE-----EE--NKPTFDS---KKKDPOVNHSQLNE 131

DB 199 GFBEQLEGKMKVDKQVVTTPPEHYAEELAGKEATFKTKVMEIKKEVPELDEIAMEL 258

QY 132 -----SHRKEDLQREHSQKSDSTKYDTATVLDKNNISSKSTTN 170

DB 259 DAEANTVDEYKENTLRRLABQKATDAMN-----EKEEATYATQDN 299

RESULT 19

PIK1 YEAST STANDARD; PRT; 1066 AA.

AC P39104;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE Phosphatidylinositol 4-kinase PIK1 (EC 2.7.1.67) (PI4-kinase)

DE (Ptdins-4-kinase).

GN PIK1 OR YNL267W OR N0795.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-JK9-3D.

RX MEDLINE=94252322; PubMed=8194527;

RA Garcia-Bustos J.F., Marini F., Stevenson I., Frei C., Hall M.N.;

RT "PIK1, an essential phosphatidylinositol 4-kinase associated with the yeast nucleus."

RT "PIK1, an essential phosphatidylinositol 4-kinase associated with the yeast nucleus."

RT Science 262:1444-1448(1993).

RL [3]

RP SEQUENCE FROM N.A.

RC STRAIN-S288c / FY1679;

RX MEDLINE=96310631; PubMed=8740425;

RA Sen-Gupta M., Lyck R., Fleig U., Niedenthal R.K., Hegemann J.H.;

RT "The sequence of a 24,152 bp segment from the left arm of chromosome XIV from Saccharomyces cerevisiae between the BNI1 and the POL2 genes."

RT Yeast 12:505-514(1996).

CC -1- FUNCTION: ACTS ON PHOSPHATIDYLINOSITOL (PI) IN THE FIRST COMMITTED STEP IN THE PRODUCTION OF THE SECOND MESSENGER INOSITOL-1,4,5-TRISPHOSPHATE. PIK1 IS PART OF A NUCLEAR PHOSPHOLIPID CYCLE AND COULD CONTROL CYTOKINESIS THROUGH THE ACTIN CYTOSKELETON.

CC -1- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-ID-myo-inositol = ADP + 1-phosphatidyl-ID-myo-inositol 4-phosphate.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- SIMILARITY: BELONGS TO THE PI3/P14-KINASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X76058; CA453658.1; -

DR EMBL; L20220; AAA34873.1; -

DR EMBL; X92494; CA463231.1; -

DR EMBL; Z71543; CA496174.1; -

DR PIR; A49335; A49335.

DR SGD; S0005211; PIK1.

DR GO; GO:0005634; C:nucleus; IDA.

DR GO; GO:0016288; P:cyclohexis; IMP.

DR GO; GO:0006892; P:post Golgi transport; IGI.

DR InterPro; IPR000403; P13\_P14\_kinase.

DR Pfam; PF00454; P13\_P14\_kinase; 1.

DR SMART; SM00146; PI3K; 1.

DR PROSITE; PS00915; P13\_4\_KINASE\_1; 1.

DR PROSITE; PS00916; P13\_4\_KINASE\_2; 1.

DR PROSITE; PS00920; P13\_4\_KINASE\_3; 1.

KW transferase; kinase; Nuclear protein.

FT DOMAIN 793 1041 PI3K/PI4K.

ST SEQUENCE 1066 AA; 119922 MW; 7665979CA1A1BC5 CRC64;

QY Query Match 10.0%; Score 89.5; DB 1; Length 1066; Best Local Similarity 22.0%; Pred. No. 43; Mismatches 59; Indels 45; Gaps 7; Matches 39; Conservative 34;

DB 4 VQDPANTTVKEPINKDTGEVSELKPHRYVTIIONGKEMSTIVSEEDFILPVYKGELE 63

DB 202 MKDFTKNTMLKNTLANKTSRSKRVSNR-----SSTPSPIDLIDPI----- 244

QY 64 KGYQDGMELISFEKKDAGYINISKDTFIRPVFKIEEKEE--ENKFTDVSKKKDN 121

DB 245 -----KTKEDASF--RKSRSSEVLDPIVDIGNOVPEEIRISSIKLPKRX 289

QY 122 PO-VNHSQNLNESHRRKEDLQREHSQKSDSTKYDTATVLDKNN--ISSKSTNNPNK 174

DB 290 PXYLDNSYHRTYDGNKINR-----DGSISNTAKALDGNKGDYISPRGRDENNE 339

RESULT 20

IGAO HAEIN STANDARD; PRT; 1694 AA.

AC P44969;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).

GN IGA OR IGA1 OR H10990.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Haemophilus.

OX NCBI\_TaxID=727;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Serotype D;
RA Wright A., Fishman Y., Tai F., Plaut A.G.;
RL Submitted (May-1991) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KM20 / ATCC 51907;
RA MEDLINE=95350630; PubMed=7542800;
RA Felschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kiehlmann A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutten G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodde A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RD Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: VIRULENCE FACTOR. CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTRACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59800; -; NOT_ANNOTATED_CDS.
DR EMBL; U32779; AAC22651.1; -.
DR PIR; H64106; H64106.
DR MEROPS; S06.001; -.
DR TIGR; H10990; -.
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertactin.
DR Pfam; PF03197; Autotransporter; 1.
DR Pfam; PF02395; IGA1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASF.
DR TIGRfam; TIGR01414; autotrans_bar1.1.
DR HydroLase; Serine protease; Transmembrane; Zymogen; signal;
KW Complete proteome.
FT SIGNAL 1 25
FT CHAIN 26 1014 POTENTIAL. IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT CONFLICT 253 254 EN -> GV (IN REF. 1).
FT CONFLICT 272 272 G -> A (IN REF. 1).
FT CONFLICT 464 464 G -> E (IN REF. 1).
FT CONFLICT 866 866 S -> T (IN REF. 1).
FT CONFLICT 1036 1036 A -> D (IN REF. 1).
FT CONFLICT 1074 1074 A -> G (IN REF. 1).
FT CONFLICT 1421 1421 A -> G (IN REF. 1).
FT CONFLICT 1545 1545 H -> T (IN REF. 1).
SQ SEQUENCE 1694 AA; 185539 MW; CS2427013F93178C CRC64;
Query Match 10.0%; Score 89.5; DB 1; Length 1694;

```

```

Best Local Similarity 23.4%; Pred. No. 71;
Matches 43; Conservative 27; Mismatches 101; Indels 13; Gaps 5;
QY 2 IVKDPFANT-----TVKFLANKDTGEVSELKPHRYVTYITONGEMSTIYSE---EDPI 54
Db 1198 VVSKQNTENTTDOPFEREKTAKVETKQF--PPQVASQAPKQOSTVOPQAVLESEN 1255
QY 55 LPVYKGELEKGYQPDGWEISGFEKGKA-GVIVINSKDTPIKPYFKIIEKKEENKPTF 113
Db 1256 VPTVNAAEVQAQLOTQTSIAVSTKQAPENSINTGSAITLTERAKSDKQOTETAASTE 1315
QY 114 DVSKKDDNPQVNSQNSHREKDLQREHS---QKSDSTQVATVLDKNNISKSTTN 170
Db 1316 DASCHKATVADNSVANNSESDPKSRRRRSISQPEISABETTAASDETTIDNDSKRS 1375
QY 171 NPK 174
Db 1376 KPNR 1379

```

```

RESULT 21
ID YCG1 YEAST STANDARD; PRT; 853 AA.
AC P25588; P25589; P27513; P87003;
DT 01-MAY-1992 (Rel. 22, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 97.9 kDa protein in CHAI-KRI intergenic region.
GN YC1061C OR YC1061C/YC106C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Rasmussen S.W., Von Wettstein D.;
RL Submitted (MAR-1992) to the EMBL/Genbank/DBJ databases.
RN [2]
RP REVISIONS.
RA Gromada R.;
RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 384-596 FROM N.A.
RC STRAIN=ATCC 28383 / FL100;
RX MEDLINE=90384830; PubMed=2169608;
RA Kern L.;
RT "The URK1 gene of Saccharomyces cerevisiae encoding uridine kinase.";
RL Nucleic Acids Res. 18:5279-5279(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59720; CAA42405.1; -.
DR EMBL; X5398; -; NOT_ANNOTATED_CDS.
DR PIR; S74279; S74279.
DR SGI; S0000566; YC1061C.
DR GO; GO:0006347; P:chromatin silencing at HML and HMR (sensu S. .); IGI.
DR GO; GO:0006348; P:chromatin silencing at telomere; IGI.
DR GO; GO:0000076; P:DNA replication checkpoint; IGI.
KW Hypothetical protein.
FT CONFLICT 505 505 L -> V (IN REF. 3).
FT CONFLICT 567 567 MISSING (IN REF. 3).
SQ SEQUENCE 853 AA; 97946 MW; 16B09FC0BF248D1 CRC64;
Query Match 9.8%; Score 88; DB 1; Length 853;
Best Local Similarity 20.6%; Pred. No. 43;
Matches 40; Conservative 30; Mismatches 80; Indels 44; Gaps 7;

```

QY 4 VKDPANTTVEKFIINKDGEVSELPKPHVTVTTONGKEMSVTSVSEEDFILPVY----- 58  
 Db 256 LRDMAEKREIVENLLEGEI-----LNKKL-----RQKRPKREKLEENFQJNANDSGSD 305  
 QY 59 KGELEGVOPDGEVIGFEG-----KQDAGVINLSKDTIKPVFKKIEK 104  
 Db 306 SGSESSGFSALSGNEIDAYESGSEENDNRRESSEKEDDEILKQKSHVYKHIINSDSD 365  
 QY 105 KEENKPTDVSKKKNPQVNSQLNESHKEDLQREHSQKSDSTKQVATVLDKNI- 163  
 Db 366 TEVEAKP-----KEKADESLEPKRIALNIGHYDNI-----GDTDKFQETNVLDTQIE 414  
 QY 164 ---SSKSTTNPNK 174  
 Db 415 EWAEERNTIENEVK 428

RESULT 22  
 FENR\_PEA STANDARD; PRT; 360 AA.  
 ID FENR\_PEA  
 AC P10933;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ferredoxin-NADP reductase, leaf isozyme, chloroplast precursor  
 DE (EC 1.18.1.2) (FNR).  
 GN PETH.  
 OS Pisum sativum (Garden pea).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.  
 OC NCBI\_TaxID=3888;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Little Marvel; TISSUE=Leaf;  
 RA Newman B.J., Gray J.C.;  
 RT "Characterisation of a full-length cDNA clone for pea ferredoxin-NADP+  
 RT reductase.";  
 RL Plant Mol. Biol. 10:511-520(1988).  
 RN [2]  
 RP SEQUENCE OF 270-360 FROM N.A., AND MUTAGENESIS.  
 RX MEDLINE=93374905; PubMed=8366077;  
 RA Orellano E.G., Calcaterra N.B., Carrillo N., Ceccarelli E.A.;  
 RT "Probing the role of the carboxyl-terminal region of ferredoxin-NADP+  
 RT reductase by site-directed mutagenesis and deletion analysis.";  
 RL J. Biol. Chem. 268:19267-19273(1993).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (1.70 ANGSTROMS).  
 RX MEDLINE=99396739; PubMed=10467097;  
 RA Deng Z., Aliverti A., Zanetti G., Arakaki A.K., Ottado J.,  
 RA Orellano E.G., Calcaterra N.B., Ceccarelli E.A., Carrillo N.,  
 RA Karpus P.A.;  
 RT "A productive NADP+ binding mode of ferredoxin-NADP+ reductase  
 RT revealed by protein engineering and crystallographic studies.";  
 RL Nat. Struct. Biol. 6:847-853(1999).  
 CC -1- FUNCTION: MAY PLAY A KEY ROLE IN REGULATING THE RELATIVE AMOUNTS  
 CC OF CYCLIC AND NON-CYCLIC ELECTRON FLOW TO MEET THE DEMANDS OF THE  
 CC PLANT FOR ATP AND REDUCING POWER.  
 CC -1- CATALYTIC ACTIVITY: Reduced ferredoxin + NADP(+) = oxidized  
 CC ferredoxin + NADPH.  
 CC -1- COFACTOR: FAD.  
 CC -1- PATHWAY: FINAL STEP IN LINEAR PHOTOSYNTHETIC ELECTRON TRANSPORT  
 CC CHAIN: IT HAS ALSO BEEN IMPLICATED IN CYCLIC ELECTRON FLOW AROUND  
 CC PHOTOSYSTEM I, ITS ROLE BEING TO RETURN ELECTRONS FROM FERREDOXIN  
 CC TO THE CYTOCHROME B-F COMPLEX.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST; STROMAL SIDE OF THE THYLAKOID  
 CC MEMBRANE IN THE VICINITY OF THE PHOTOSYSTEM I IN THE NON-STACKED  
 CC AND FRINGE PORTION OF THE MEMBRANE.  
 CC -1- MISCELLANEOUS: FNR IS PROBABLY ATTACHED TO THE MEMBRANE BY A  
 CC SPECIFIC BINDING PROTEIN.  
 CC -1- SIMILARITY: WITH OTHER SPECIES FNR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X12446; CAA30978.1; -;  
 CC EMBL; L15565; AAB59349.1; -;  
 CC EMBL; L15567; AAB59303.1; -;  
 CC EMBL; L15569; AAB59304.1; -;  
 CC PIR; S04030; S04030.  
 CC PDB; 1OFZ; 01-SEP-99.  
 CC PDB; 1OFY; 01-SEP-99.  
 CC PDB; 1OG0; 01-SEP-99.  
 CC PDB; 1OGA; 01-SEP-99.  
 CC InterPro; IPR001709; FPN\_cyt\_redctse.  
 CC InterPro; IPR001433; Oxred\_FAD/NAD(P).  
 CC Pfam; PF00175; NAD\_binding\_1; 1.  
 CC PRINTS; PR00371; FPNCR.  
 CC Oxidoreductase; Flavoprotein; NADP; FAD; Chloroplast; Transit peptide;  
 CC Electron transport; Photosynthesis; Thylakoid; Membrane;  
 CC MultiGene family; 3D-structure.  
 CC TRANSIT 1 52  
 CC CHAIN 53 360  
 CC DISULFID 178 183  
 CC NP BIND 212 230  
 CC MUTAGEN 360 360  
 CC STRAND 68 68  
 CC TURN 71 72  
 CC STRAND 78 78  
 CC TURN 80 81  
 CC STRAND 84 93  
 CC TURN 97 98  
 CC STRAND 103 109  
 CC TURN 111 112  
 CC STRAND 118 119  
 CC TURN 121 125  
 CC STRAND 129 129  
 CC TURN 131 132  
 CC STRAND 135 135  
 CC TURN 139 143  
 CC STRAND 147 148  
 CC TURN 150 151  
 CC STRAND 156 162  
 CC TURN 165 167  
 CC STRAND 169 170  
 CC TURN 173 175  
 CC STRAND 177 184  
 CC TURN 187 188  
 CC STRAND 190 197  
 CC TURN 200 201  
 CC STRAND 204 204  
 CC TURN 207 208  
 CC STRAND 210 216  
 CC TURN 217 220  
 CC STRAND 221 221  
 CC TURN 222 222  
 CC STRAND 223 222  
 CC TURN 224 225  
 CC STRAND 225 250  
 CC TURN 253 255  
 CC STRAND 257 258  
 CC TURN 259 268  
 CC STRAND 270 272  
 CC TURN 272 272  
 CC STRAND 273 279  
 CC TURN 280 282  
 CC STRAND 284 284  
 CC TURN 286 287  
 CC STRAND 290 290  
 CC TURN 293 297  
 CC STRAND 298 300  
 CC TURN 301 309  
 CC STRAND 311 312  
 CC TURN 312 312

```
FT STRAND 313 319
FT TURN 321 322
FT HELIX 323 337
FT TURN 338 339
FT HELIX 342 351
FT TURN 352 353
FT STRAND 355 360
SQ SEQUENCE 360 AA; 40194 MW; 7F1CC10DEBBA7B24 CRC64;

Query Match 9.8%; Score 87.5; DB 1; Length 360;
Best Local Similarity 24.6%; Pred. No. 19;
Matches 41; Conservative 21; Mismatches 54; Indels 51; Gaps 8;

Qy 4 VKDFARNTT---VKEFTLNKDTGEV-----SELKP-HRVTVITQNGKEM----- 43
Db 148 IGDGDSKTVSLVCKRLVYTTDAGEVVGKVCNSFLCDLKPGEVKITGPGVKEMLMKPKDP 207
Qy 44 -----STVISEEDFILPVYKGELEKGYQDFD---WEISGPEGKDAAGYVNLKSDT 92
Db 208 NATVIMLTGTGCIAPFRSFLWKF-FEKHEDYQFNLAWLFLG-----VPTSSL 256
Qy 93 FIKPVFKIEKKKEENKPTDVSKKDNPQVNHSHRKHEDLQ 139
Db 257 LYKEFEKMKKAPENFRDFAVSRE-----QVNDXGKMYIQ 294

RESULT 23
SKM1_YEAST
ID SKM1_YEAST STANDARD; PRT; 655 AA.
AC Q12469; Q06940;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine/threonine-protein kinase SKM1 (EC 2.7.1.-) (Protein kinase
75490 D).
GN SKM1 OR YOL113W OR HRA655.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / GRF88;
RX MEDLINE=97197189; PubMed=9042478;
RA Martin H., Mendoza A., Rodriguez-Pachon J.M., Molina M., Nombela C.;
RT "Characterization of SKM1, a Saccharomyces cerevisiae gene encoding a
novel Ste20/PAK-like protein kinase."
RL Mol. Microbiol. 23:431-444 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96076631; PubMed=7502582;
RA Vandenbol M., Durand P., Portetelle D., Hilger F.;
RT "Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV
including the Ty1-H3 retrotransposon, the sufl(+) frameshift
suppressor gene for trna-Gly, the yeast transfer RNA-Thr-1a and a
delta element."
RL Yeast 11:1069-1075 (1995).
CC -!- FUNCTION: MAY BE INVOLVED IN CELLULAR SIGNALLING OR CYTOSKELETAL
CC FUNCTIONS. MAY PLAY A ROLE IN MORPHOGENETIC CONTROL.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC STE20 SUBFAMILY.
CC -!- SIMILARITY: Contains 1 CRIB domain.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X69322; CAA49163.1; -.
```

```
DR EMBL; Z49149; CAA88147.1; -.
DR EMBL; Z74855; CAA99132.1; -.
DR PIR; S51884; S51884.
DR SGD; S0005473; SKM1.
DR GO; GO:0000074; P:regulation of cell cycle; IGI.
DR InterPro; IPR000095; PAKbox/RhoBindng.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_Thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00786; PBD; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00285; PBD; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS50108; CRIB; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR Transferase; Serine/threonine-protein kinase; ATP-binding.
KW DOMAIN 3 118
FT DOMAIN 123 136
FT DOMAIN 360 639
FT NP_BIND 366 374
FT BINDING 406 406
FT ACT_SITE 507 507
FT CONFLICT 303 303
FT CONFLICT 306 306
FT CONFLICT 320 320
FT CONFLICT 322 322
FT CONFLICT 452 452
FT CONFLICT 468 468
SQ SEQUENCE 655 AA; 75331 MW; 1F6CBC85FE10D385 CRC64;

Query Match 9.8%; Score 87.5; DB 1; Length 655;
Best Local Similarity 20.5%; Pred. No. 35;
Matches 34; Conservative 36; Mismatches 57; Indels 39; Gaps 7;

Qy 16 FILNKDTGEVSELKPHRVTVITQNGKEMSTVISEDFILPVYKGE-----EKGQYOF 68
Db 209 YIINKRTNSIK---RSVSRTLRKGK-----TDSILPVYQSELKPPRPSPDDYKF 255
Qy 69 DGEWISGF--EGKXDAGYVNLKSDTIFKPVFKIEKKEE-----ENKPTFDVSK 117
Db 256 TNIEDNKVREGR-----VHVSKESTADSQTKLGKKEQKVQSHLRRHNNSTFRPHR 309
Qy 118 KKDNPQV--NHSQNLNESHKEDLQREHSHQKSDSTKDVATVLDKN 161
Db 310 LAPSAPATKHSDSKTKWHKEDLLELKNDDSNELIMKMTVAIDVN 355

RESULT 24
GARP_PLAFF STANDARD; PRT; 678 AA.
AC P13816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glutamic acid-rich protein precursor.
GN GARP.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5937;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89040048; PubMed=2903445;
RA Triglia T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA Kemp D.J.;
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
acid-rich protein (GARP).";
```

```
RL Mol. Biochem. Parasitol. 31:199-202(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03998; AAA29605.1; -
DR PIR; A54514; A54514.
KW Repeat; Malaria; Antigen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 678 GLUTAMIC ACID-RICH PROTEIN.
FT DOMAIN 120 164 15 X 3 AA TANDEM REPEATS OF K-K-X.
FT DOMAIN 372 416 9 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 417 441 5 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 576 604 POLY-GLU
FT DOMAIN 605 653 7 X APPROXIMATE TANDEM REPEATS.
FT DOMAIN 654 663 POLY-GLU.
SQ SEQUENCE 678 AA; 80551 MW; 2A8F85606496EA9E CRC64;

Query Match 9.8%; Score 87.5; DB 1; Length 678;
Best Local Similarity 24.1%; Pred. No. 37; Mismatches 73; Indels 37; Gaps 10;
Matches 46; Conservative 35;

QY 5 KDFARNTTVKEFILNKD-----TGEV---SELKPHRVTVTIQNGKMSSTIVSEE-----DF 53
DB 25 KCFNSGLLNQNILNKSPFISITGRLLNETELEKNK-----DINSKSTLLKKEKDKDD 78
QY 54 ILPVYGELEKGVQPCWEISGFEKKGADGVINLSKDTFKIPVKFKEKK-----EE 107
DB 79 VPTTSNDNLKNAHNN--BIS---SSTDPTNIINVNDKDNENSVDKKKKKKKKKK 133
QY 108 ENKPTFDVSKKDNPOVNSHLSNESHKEDLOREHSQ-----KSDSTKDTVATVLDKNN 162
DB 134 EKKEKKDKKKKKKKKKKK--EKKKKKKKKKKENSEVMSLYKTKGQHPKPKATEHGEEN 191
QY 163 ISSK--STNN 171
DB 192 LDEMYVSEINN 202

RESULT 25
RA50 PYRHO STANDARD; PRT; 879 AA.
AC O58687:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR PH0929.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3.
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
```

```
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -1- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000004; BAA30025.1; -
DR PIR; C71083; C71083.
DR HAMAP; MF_00449; -; 1.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF04423; Rad50 zn hook; 1.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR SMART; SM00382; AAA; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 AIP (BY SIMILARITY).
FT DOMAIN 141 744 COILED COIL (POTENTIAL).
SQ SEQUENCE 879 AA; 103673 MW; FD4E30FD1BBCB29 CRC64;

Query Match 9.8%; Score 87.5; DB 1; Length 879;
Best Local Similarity 21.0%; Pred. No. 49; Mismatches 51; Indels 27; Gaps 3;
Matches 30; Conservative 35;

QY 4 VKDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELE 63
DB 551 LNDYKNSTKLEIETDKAKKELSEIEDRLRLGFKTIDELSGRI-----RELE 598
QY 64 KYQFDGWEISGFEKKGADGVINLSKDTFKIPVKFKEKKKEENKTFDFVSKKKNPQ 123
DB 599 K-----PHNK---YIEAKNAEKELRDILESUKDERELDKAPEELAKIETDIE 643
QY 124 VNHSQLNESHKEDLOREHSOK 146
DB 644 KVTSQLNELQRFQKKYEEKRE 666

RESULT 26
UN89 CAEEL STANDARD; PRT; 6632 AA.
AC O01761; Q17362;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09pl.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
RT assembly, encodes a giant modular protein composed of Ig and signal
RT transduction domains."
RL J. Cell Biol. 132:835-848(1996).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Le T.T., Wilson R.;
```

```

RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofibrilament
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBI-homology (DH) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCSD domains.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U33058; AAB00542.1; -.
CC EMBL; AF003131; AAB54132.2; -.
CC PDB; 1PHO; 20-DEC-00.
CC WormPep; C09D1.1; CE30426.
CC InterPro; IPR003961; FN III.
CC InterPro; IPR007110; IG-like.
CC InterPro; IPR003598; IG C2.
CC InterPro; IPR003306; IG_MHC.
CC InterPro; IPR001849; PH.
CC InterPro; IPR007850; RCSD.
CC InterPro; IPR000219; RhoGEF.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00041; fn3; 1.
CC Pfam; PF00047; ig; 47.
CC Pfam; PF00165; PH; 1.
CC Pfam; PF01577; RCSD; 5.
CC Pfam; PF00621; RhoGEF; 1.
CC Pfam; PF00018; SH3; 1.
CC SMART; SM00408; IGc2; 23.
CC SMART; SM00325; RhoGEF; 1.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS50010; DH 2; 1.
CC PROSITE; PS50835; IG Like; 49.
CC PROSITE; PS50003; PH DOMAIN; 1.
CC PROSITE; PS50002; SH3; 1.
KW Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KW 3D-structure.
FT DOMAIN 63 127 SH3.
FT DOMAIN 152 330 DH.
FT DOMAIN 342 498 PH.
FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT DOMAIN 548 736 IG-LIKE C2-TYPE 2.
FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
FT DOMAIN 1272 1315 THR-RICH.
FT DOMAIN 1375 1475 RCSD 1.
FT DOMAIN 1479 1585 RCSD 2.
FT DOMAIN 1597 1695 RCSD 3.
FT DOMAIN 1700 1799 RCSD 4.
FT DOMAIN 1800 1860 RCSD 5.
FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.

```

Query Match 9.7%; Score 87; DB 1; Length 6632;  
 Best Local Similarity 21.8%; Pred. No. 4.7e+02;





CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; M88098; AAA29744.1; -;  
 KW Malaria; Receptor; Membrane.  
 FT NON\_TER 1251 1251  
 SQ SEQUENCE 1251 AA; 143741 MW; 54BA51C7404AC572 CRC64;

Query Match 9.6%; Score 86.5; DB 1; Length 1251;  
 Best Local Similarity 22.8%; Pred. No. 84;  
 Matches 46; Conservative 40; Mismatches 77; Indels 39; Gaps 10;

QY 2 IWVKDFARNTTVKPF-----ILNKD---TGEVSELKPH--RVTVTTIQNGKEMSTI--- 47  
 DB 234 IALEDAQIDTEVSKIEQINREIMNKDKBIKSYLSEIKYKDKCTTEISNSKRGDKIEFL 293  
 QY 48 -----VSEHDFILPVYKGELEKGYQFDGWEISGFE-GKKDAGYVNL--SKDTFKIPVFKK 100  
 DB 294 EKFPNEESNSKVNINEINIRNSEQYLKDIEDAEKQASTKVELFKHKHETTISNIFKE 353  
 QY 101 -----IEBKKEENKPTDVSK--KKDNPOV-----NHSQLNESHKREDLQREHS 144  
 DB 354 SEILGVETYSOKKINKAEDIMKEIERHNSHSEIQTVKGFQENLNKLNPHNYDNADELNN 413  
 QY 145 QKSDSTKDVATVLD--KNNIS 164  
 DB 414 DKSTNAKVLIEITNLESVKHNL 435

# RESULT 29

SYST LYES  
 ID SYST LYES STANDARD; PRT; 200 AA.  
 AC P27058;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Systemin precursor.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4081;  
 RN [1]  
 RC TISSUE=Leaf.  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92196587; PubMed=1549783;  
 RA McGurl B., Pearce G., Orozco-Cardenas M., Ryan C.A.;  
 RT "Structure, expression, and antisense inhibition of the systemin precursor gene."  
 RL Science 255:1570-1573 (1992).  
 RN [2]  
 RP SEQUENCE OF 179-196.  
 RA Pearce G., Strydom D., Johnson S., Ryan C.A.;  
 RT "A polypeptide from tomato leaves induces wound-inducible proteinase inhibitor proteins."  
 RL Science 253:895-898 (1991).  
 RN [3]  
 RP FUNCTION.  
 RX MEDLINE=20175266; PubMed=10708853;  
 RA Ryan C.A.;  
 RT "The systemin signaling pathway: differential activation of plant defensive genes";  
 RL Biochim. Biophys. Acta 1477:112-121 (2000).  
 CC -!- FUNCTION: ACTIVATES A LIPID-BASED SIGNAL TRANSDUCTION PATHWAY IN WHICH LINOLENIC ACID IS CONVERTED TO JASMONIC ACID, A POTENT ACTIVATOR OF DEFENSE GENE TRANSCRIPTION, INCLUDING PROTEINASE INHIBITOR.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- TISSUE SPECIFICITY: ALL ORGANS EXCEPT THE ROOTS. TRANSPORTED OUT OF WOUNDS TO DISTAL TISSUES.  
 CC -!- INDUCTION: By wounding; in leaves.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; M84800; AAA34182.1; -;  
 DR EMBL; M84801; AAA34184.1; -;  
 DR PIR; T07149; T07149.  
 KW Hormone; Repeat.

FT PROPEP 1 178  
 FT CHAIN 179 196  
 FT PROPEP 197 200  
 FT REPEAT 3 8  
 FT REPEAT 37 45  
 FT REPEAT 80 88  
 FT REPEAT 117 125  
 FT REPEAT 145 153  
 SQ SEQUENCE 200 AA; 22999 MW; FA251D94BAA9CSA9 CRC64;

Query Match 9.6%; Score 86; DB 1; Length 200;  
 Best Local Similarity 21.5%; Pred. No. 13;  
 Matches 37; Conservative 30; Mismatches 59; Indels 46; Gaps 8;

QY 1 KIVVK-----DFARNTTVKGFILNKDTGEVSELKPHRVTVTTIQNGKEMSTIIVERDIL 55  
 DB 34 KIEKETPSQDINNKTISYVLDRTDQEIPOE-H-----EGGYVKEKIVEKETISQ 86  
 QY 56 PVYKGELEKGYQFDGWEISGFEKGDAGYVNLNISKDTFKIPVFKKIEEKE---EENKPT 112  
 DB 87 YIIK-----IEGDDA-----QEKLVYEVEEVEEKEKIVEKETPS 122  
 QY 113 FDYKSKKDN-----POVNHSQLNESHKREDLQREHSQKSDSKDVTATVLDK 160  
 DB 123 QDINNKGDDAQEKPKVEH---EGDDKETSQDIIRMEGEGALEITKVCKE 171

# RESULT 30

GLNA\_PIRKO  
 ID GLNA\_PIRKO STANDARD; PRT; 443 AA.  
 AC O08467;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).  
 GN GLNA.  
 OS Pyrococcus kodakaraensis.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae; Thermococcaceae;  
 OX NCBI\_TaxID=69014;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KOD1;  
 RX MEDLINE=97316461; PubMed=9172372;  
 RA Adul Rahman R.N.Z., Jongsareejit B., Fujiwara S., Imanaka T.;  
 RT "Characterization of recombinant glutamine synthetase from the hyperthermophilic archaeon Pyrococcus sp. strain KOD1."  
 RL Appl. Environ. Microbiol. 63:2472-2476 (1997).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate + L-glutamine.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE GLUTAMINE SYNTHETASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; D86222; BAA20530.1; --  
 DR HSP; P06201; ILGR.  
 DR InterPro; IPR001691; GLN synth.  
 DR InterPro; IPR004809; GLN.  
 DR InterPro; IPR001637; GlnA adenyltn.  
 DR Pfam; PF00120; gln-synt\_1.  
 DR Pfam; PF03951; gln-synt\_N; 1.  
 DR ProDom; PD001057; Gln synt\_C; 1.  
 DR TIGRFAMs; TIGR00653; GlnA; 1.  
 DR PROSITE; PS00180; GLNA 1; 1.  
 DR PROSITE; PS00181; GLNA\_ATP; 1.  
 KW Ligase  
 FT BINDING 362 362 AMP (UNDER CONDITIONS OF ABUNDANT  
 FT GLUTAMINE) (BY SIMILARITY).  
 SQ SEQUENCE 443 AA; 50259 MW; 9426DCCFEF18168 CRC64;  
 Query Match 9.6%; Score 86; DB 1; Length 443;  
 Best Local Similarity 17.8%; Pred. No. 30;  
 Matches 31; Conservative 25; Mismatches 46; Indels 72; Gaps 4;  
 QY 58 YKGELEKGYQDFGWEISGFEKGDAGYVINSKDTFI----- 94  
 DB 42 YEAEVDGVSFDGSSIPGEGIEDSLIFKADPSTVAEIPWEGIGRVGYIKGDEPYQA 101  
 QY 95 --KPVFKKTEERKEEN-----KPTFDVSKKD----- 120  
 DB 102 DPGILKRVLERLEKGLKAHIGPEPEFYIFKNGTWELHPDGSYFDVLGLDKAREIR 161  
 QY 121 -----NPOVNSQLNESHRKEDLQREHSOKSD---STKDVATATVLD 159  
 DB 162 REIALYMPYLGLKPEVLHHEVKAQHEIDFRIDEALRTADNIVSKHVKAQAE 215  
 RESULT 31  
 ID PBPA\_BACSU STANDARD; PRT; 914 AA.  
 AC P39793;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Penicillin-binding protein 1A/1B (PBPI) [includes: Penicillin-  
 DE insensitive transglycosylase (EC 2.4.2.-) (peptidoglycan TGase);  
 DE Penicillin-sensitive transpeptidase (EC 3.4.-.-) (DD-transpeptidase)].  
 GN PONA.  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]\_TaxID=1423;  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 499-515.  
 RC STRAIN=168;  
 RX MEDLINE=95113769; PubMed=7814321;  
 RA Popham D.L., Setlow P.;  
 RT "Cloning, nucleotide sequence, and mutagenesis of the Bacillus  
 RT subtilis pona operon, which codes for penicillin-binding protein  
 RT (PBP) 1 and a PBP-related factor."  
 RL J. Bacteriol. 177:326-335 (1995).  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=168 / Marburg;  
 RC STRAIN=168 / Marburg;  
 RX MEDLINE=96349105; PubMed=8760912;  
 RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,  
 RA Serror P.;  
 RT "Sequence analysis of the Bacillus subtilis chromosome region between  
 RT the serA and kgd loci cloned in a yeast artificial chromosome."  
 RL Microbiology 142:2005-2016 (1996).  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=168;  
 RC MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,

RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
 RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portecelle D., Forwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
 RT subtilis.";  
 RN Nature 390:249-256 (1997).  
 RN [4]  
 RP GROWTH REQUIREMENTS.  
 RC STRAIN=168;  
 RX MEDLINE=98389671; PubMed=9721295;  
 RA Murray T., Popham D.L., Setlow P.;  
 RT "Bacillus subtilis cells lacking penicillin-binding protein 1 require  
 RT increased levels of divalent cations for growth."  
 RL J. Bacteriol. 180:4555-4563 (1998).  
 RN [5]  
 RP SUBCELLULAR LOCATION.  
 RC STRAIN=168;  
 RX MEDLINE=99255546; PubMed=10322023;  
 RA Pedersen L.B., Argert E.R., Setlow P.;  
 RT "Septal localization of penicillin-binding protein 1 in Bacillus  
 RT subtilis.";  
 RL J. Bacteriol. 181:3201-3211 (1999).  
 CC -!- FUNCTION: CELL WALL FORMATION. SYNTHESIS OF CROSS-LINKED  
 CC PEPTIDOGLYCAN FROM THE LIPID INTERMEDIATES. THE ENZYME HAS A  
 CC PENICILLIN-INSENSITIVE TRANSGLYCOSYLASE N-TERMINAL DOMAIN  
 CC (FORMATION OF LINEAR GLYCAN STRANDS) AND A PENICILLIN-SENSITIVE  
 CC TRANSEPTIDASE C-TERMINAL DOMAIN (CROSS-LINKING OF THE PEPTIDE  
 CC SUBUNITS) (BY SIMILARITY).  
 CC -!- PATHWAY: Peptidoglycan synthesis; final stages.  
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, PROBABLY FOUND ALL  
 CC OVER THE WHOLE CELL AT LOW CONCENTRATIONS. ALSO LOCALIZES TO THE  
 CC DIVISION SITE IN VEGETATIVE CELLS.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSION IS CONSTANT DURING GROWTH,  
 CC DECREASES DURING SPORULATION AND IS INDUCED APPROXIMATELY 15 MIN  
 CC INTO SPORE GERMINATION.  
 CC -!- PTM: THE PRODUCT EXPRESSED FROM THE TRANSLATION OF THE PONA GENE  
 CC APPEARS AS TWO BANDS ON A GEL (1A AND 1B), BUT THE SPECIFIC AMINO  
 CC ACID SEQUENCE OF EACH PROTEIN IS UNKNOWN.  
 CC -!- PTM: THE N-TERMINUS IS BLOCKED.  
 CC -!- MISCELLANEOUS: CELLS LACKING THE PROTEIN REQUIRE INCREASED LEVELS  
 CC OF MG(2+) OR CA(2+) FOR GROWTH AND GERMINATION. APPROXIMATELY 50%  
 CC OF CELLS WITHOUT THE PROTEIN CONTAIN ABNORMAL FTZ RINGS,  
 CC SUGGESTING IT IS INVOLVED IN SEPTUM SYNTHESIS; INCREASED LEVELS OF  
 CC MG(2+) OR CA(2+) ONLY PARTIALLY ELIMINATE THE SEPTATION DEFECTS.  
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE  
 CC TRANSGLYCOSYLASE FAMILY.  
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE  
 CC TRANSEPTIDASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



```

QY 142 EHSQSDS 149
DB 4747 EDEKSDS 4754

RESULT 33
DRS1_YEAST
ID DRS1_YEAST STANDARD; PRT; 752 AA.
AC P32892;
DT 01-OCT-1993 (Rel. 27, Last Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable ATP-dependent RNA helicase DRS1.
GN DRS1 OR YLI008W OR L1345.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93087480; PubMed=1454790;
RA Ripmaster T.L., Vaughn G.P., Woolford J.L. Jr.;
RT "A putative ATP-dependent RNA helicase involved in Saccharomycetes
RT cerevisiae ribosome assembly."
RL Proc. Natl. Acad. Sci. U.S.A. 89:11131-11135 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / FY23;
RX MEDLINE=96405918; PubMed=8810043;
RA Miosga T., Zimmermann F.K.;
RT "Sequence analysis of the CEN12 region of Saccharomyces cerevisiae on
RT a 43.7 kb fragment of chromosome XII including an open reading frame
RT homologous to the human cystic fibrosis transmembrane conductance
RT regulator protein CFTR."
RL Yeast 12:693-708 (1996).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / AB972;
RX MEDLINE=97313267; PubMed=9169871;
RA Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W.,
RA Benes V., Brueckner M., Delius H., Dubois E., Duesterhoeft A.,
RA Entian K.-D., Floeth M., Goffeau A., Hebling U., Heumann K.,
RA Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Koetter P.,
RA Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Moestl D.,
RA Mueller-Auer S., Newnich U., Obermaier B., Piravandi E., Pohl T.M.,
RA Portetelle D., Purnelle B., Reckmann S., Rieger M., Rinke M., Rose M.,
RA Scharte M., Scherens B., Scholler P., Schwager C., Schwarz S.,
RA Underwood A.P., Urrestazu L.A., Vandenbol M., Verhasselt P.,
RA Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E.,
RA Wedler H., Zimmermann F.K., Zollner A., Hani J., Hohelsel J.D.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII."
RL Nature 387:87-90 (1997).
CC -!- FUNCTION: PROBABLE HELICASE INVOLVED IN RIBOSOME ASSEMBLY.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L00683; AAA34666.1; -
CC EMBL; X21488; CAA62783.1; -
CC EMBL; Z73113; CAA97452.1; -
CC FIP; S64750; S64750.
CC HSP; Q58083; 1HV8.
CC SGD; S0003931; DRS1.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR000629; DEAD_box.
CC InterPro; IPR001650; Helicase_C.

DR Pfam: PF00270; DEAD; 1.
DR Pfam: PF00271; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICc; 1.
DR PROSITE; PS00039; DEAD ATP HELICASE; 1.
KW ATP-binding; RNA-binding; Helicase; Nuclear protein.
FT DOMAIN 170 190 POLY-GLU.
FT NP_BIND 275 282 ATP (POTENTIAL).
FT SITE 385 388 DEAD BOX.
FT CONFLICT 1 68
FT
SQ SEQUENCE 752 AA; 84843 MW; 60747607A6E5E4A8 CRC64;
Query Match 9.5%; Score 85.5; DB 1; Length 752;
Best Local Similarity 21.8%; Pred. No. 57;
Matches 38; Conservative 29; Mismatches 64; Indels 43; Gaps 8;
QY 17 ILNKDTGEVSELKPHRVTVTTIQNGKMSSTTVSEEDFILPVYKGEKGYQFD----- 69
DB 26 ILDSDDDEKVEAKK---TKRKKGKN-NKKVSEGNLDEVDHEDLDAGKFDLDADDTT 81
QY 70 ----GWEI--SGFEKGDAGVYVNLKDTFKVPFK-----IBEKKEENKPTFD 114
DB 82 SNFGQWNLAEAGESNKKDAEAFV--KKDVLDKIIRKGGGLVKMAHIDSKQEETEKE-K 138
QY 115 VSKKKDN-----PQVNHSQLNSHRKEDLQREHHSOKSDSTKDV 154
DB 139 VEKENDSDDEALMDGFGMGAPMNGDENQSEEEEEKEEEEEEEQEEEMT 192

RESULT 34
RGAL_YEAST
ID RGAL_YEAST STANDARD; PRT; 1007 AA.
AC P39083; P39934;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Rho-type GTPase-activating protein 1.
GN RGAL OR DBM1 OR THE1 OR YOR127W OR O3290 OR YOR3290W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RX STRAIN=S288c;
RX MEDLINE=96239492; PubMed=8657111;
RA Chen G.-C., Zheng L., Chan C.S.M.;
RT "The LIM domain-containing Dbp1 GTPase-activating protein is required
RT for normal cellular morphogenesis in Saccharomyces cerevisiae."
RL Mol. Cell. Biol. 16:1376-1390 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / FY1679;
RX MEDLINE=97060020; PubMed=8904341;
RA Wienann S., Reckmann S., Benes V., Voss H., Schwager C., Vlcek C.,
RA Stegemann J., Zimmermann J., Exfle H., Paces V., Ansorge W.;
RT "Sequencing and analysis of 51 kb on the right arm of chromosome XV
RT from Saccharomyces cerevisiae reveals 30 open reading frames."
RL Yeast 12:281-288 (1996).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97344368; PubMed=9200815;
RA Voss H., Benes V., Andrade M.A., Valencia A., Reckmann S., Teodoru C.,
RA Schwager C., Paces V., Sander C., Ansorge W.;
RT "DNA sequencing and analysis of 130 kb from yeast chromosome XV."
RL Yeast 13:655-672 (1997).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c;
RX MEDLINE=96101594; PubMed=7498791;
RA Stevenson B.J., Ferguson B., de Virgilio C., Bi E., Pringle J.R.,

```







RESULT 38



DT 01-OCT-1996 (Rel. 34, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein C106.14c in chromosome II.  
GN SPBC106.14C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Voickaert G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vansteelt E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritsch C., Leitz C., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Cloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe."  
RL Nature 415:871-880 (2002).  
CC -1- SIMILARITY: TO YEAST YGR245C.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AL10295; CAB53730.1; -;  
DR F1R; T39271; T39271.  
DR GeneDB SPombe; SPBC106.14c; -;  
DR Pfam; PF05285; SDAL; 1.  
KW Hypothetical protein.  
FT DOMAIN 250 254 POLY-GLU.  
FT DOMAIN 559 562 POLY-GLU.  
SQ SEQUENCE 719 AA; 81250 MW; 93864D55B86BD1D6 CRC64;  
  
Query Match 9.5%; Score 85; DB 1; Length 719;  
Best Local Similarity 23.9%; Pred. No. 59;  
Matches 34; Conservative 32; Mismatches 62; Indels 14; Gaps 5;  
  
QY 15 EFLNKDGTGEVSELKPHRVTVITQNGKMSSTIVSEDFILPYKGEI-----EKGYQFDG 70  
Db DMLKRRDGRKGLASIEKMDRT-PLKYGEELNVTHGIQGLELLAQYKAEGHEGEGDWDN 527  
QY 71 WEIS-GFEKKDAGYVINLSKDTFKIPVFKTEEKKEENKPTFDVSKKDKDPQVHSQL 129  
Db WEVSEDSQNSDDSGGWIDVSDNNI-----ELSDSDEEEKAT---ARKESDEKSSQK 579  
QY 130 NESHKREDLQREHSQKSDSTK 151  
  
Db 580 ELVDRMTELASQSIILTPNDLKK 601  
  
RESULT 40  
ID IF2\_STAEP STANDARD; PRT; 720 AA.  
AC Q8CST4;  
DT 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DE Translation initiation factor IF-2.  
DE Translation initiation factor IF-2.  
GN INF2 OR SE0945.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
RA Chen Z., Wen Y.;  
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
RL FUNCTION: One of the essential components for the initiation of  
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous  
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.  
CC Also involved in the hydrolysis of GTP during the formation of the  
CC 70S ribosomal complex (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE016747; AA004542.1; -;  
DR HAMAP; MF\_00100; -; 1.  
DR Pfam; PF00009; GTP\_EFTU; 1.  
DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
DR Pfam; PF04760; IF2\_N; 2.  
DR ProDom; PD186100; IF2; 1.  
DR TIGRFAMs; TIGR00487; IF-2; 1.  
DR TIGRFAMs; TIGR00231; small\_GTP; 1.  
DR PROSITE; PS01176; IF2; 1.  
KW Initiation factor; Protein biosynthesis; GTP-binding;  
KW Complete proteome.  
FT DOMAIN 225 373 G-DOMAIN.  
FT NP\_BIND 231 238 GTP (BY SIMILARITY).  
FT NP\_BIND 277 281 GTP (BY SIMILARITY).  
FT NP\_BIND 331 334 GTP (BY SIMILARITY).  
SQ SEQUENCE 720 AA; 79343 MW; 07FBSA6A59CF970C CRC64;  
  
Query Match 9.5%; Score 85; DB 1; Length 720;  
Best Local Similarity 28.0%; Pred. No. 59;  
Matches 23; Conservative 15; Mismatches 36; Indels 8; Gaps 2;  
  
QY 101 IBEKKEENKPTFDVSKKD-----NPQVNHSQLNESHKEDLQREHSQKSDSTK----D 152  
Db LEEEQIKALDKKFKASQAKDTNKQNTQNNHQSNNKQNSNDKQKQSKNNKPTKKQCN 97  
QY 153 VTATVLDKNNISSKSTTNPNK 174  
Db NKGKQNNKNNKNNKNNKNNK 119  
  
Search completed: February 10, 2004, 10:49:56  
Job time : 9.08902 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 35.0346 Seconds  
(without alignments)  
697.707 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_620\_773

Perfect score: 799

Sequence: 1 DTGVSSELKPHRVTTIQNG.....ATVLDKNNISKSTTNNPNK 154

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2001.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2002.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	799	100.0	773	22	AAB48343
2	799	100.0	2120	21	AA181710
3	799	100.0	2140	24	ABU01020
4	615	77.0	117	19	AAW55096
5	615	77.0	117	23	ABP54590
6	118	14.8	746	22	AA181779
7	118	14.8	778	23	ABP39023
8	107.5	13.5	484	21	AA181777
9	105	13.1	665	21	AA1818278

10	101.5	12.7	564	22	ABB61977	Drosophila melanog
11	96.5	12.1	2060	23	AAE20967	Staphylococcus lug
12	96	12.0	2500	21	AA1818272	Plasmodium falcipa
13	93.5	11.7	408	21	AAG371134	Arabidopsis thalia
14	93.5	11.7	456	21	AAG371133	Arabidopsis thalia
15	93.5	11.7	476	21	AAG371132	Arabidopsis thalia
16	92.5	11.6	2519	22	ABG16636	Novel human diagn
17	90.5	11.3	209	21	AA1818277	Human ORFX ORF2661
18	90.5	11.3	757	21	AA1818278	Human cell cycle r
19	90.5	11.3	758	21	AA1818279	Human colon cancer
20	90	11.3	442	23	ABP38188	Staphylococcus epi
21	90	11.3	645	24	ABP38189	Pathogen specific
22	90	11.3	654	24	ABP38190	Staphylococcus epi
23	90	11.3	1183	22	ABP38191	Drosophila melanog
24	89.5	11.2	402	20	AA1818280	S. aureus trigger
25	89.5	11.2	511	20	AA1818281	Chlamydia pneumoni
26	89.5	11.2	525	20	AA1818282	S. aureus trigger
27	89.5	11.2	1408	22	ABP38192	Drosophila melanog
28	89	11.1	645	24	ABP38193	Pathogen specific
29	88.5	11.1	281	22	ABP38194	Drosophila melanog
30	88.5	11.1	1141	22	AA1818283	Shrimp white spot
31	88	11.0	225	23	ABP38195	Candida albicans e
32	88	11.0	258	22	AA1818284	Human protein sequ
33	88	11.0	817	22	AA1818285	Human protein SQ
34	88	11.0	817	22	AA1818286	Human protein SQ
35	87.5	11.0	635	23	ABP38196	Candida albicans e
36	87.5	11.0	2515	22	ABP38197	Human secreted pro
37	87.5	11.0	2515	22	AA1818287	Human protein SQ
38	87.5	11.0	2515	22	AA1818288	Human protein SQ
39	87.5	11.0	3021	24	ABP38198	Human secretory po
40	87	10.9	313	21	AA1818289	Arabidopsis thalia
41	87	10.9	607	22	AA1818290	Human polypeptide
42	87	10.9	635	21	AA1818291	Plasmodium falcipa
43	87	10.9	902	23	ABP40312	Staphylococcus epi
44	87	10.9	922	22	AA1818292	Shrimp white spot
45	86.5	10.8	443	22	AA1818293	Salmonella typhi c

ALIGNMENTS

RESULT 1

AAB48343  
ID AAB48343 standard; Protein; 773 AA.

XX AAB48343;

AC AAB48343;

XX 20-APR-2001 (first entry)

DT S. pneumoniae Spi30 polypeptide.

DE Immunogenic; Spi28; Spi30; pneumococcal; otitis media; nasopharyngeal;

KW bronchial; lung; blood; infection; immune response; immunotherapy;

KW antibacterial; auditory; vaccine.

XX Streptococcus pneumoniae.

OS Streptococcus pneumoniae.

XX WO2000076540-A2.

XX 21-DEC-2000.

PF 09-JUN-2000; 2000WO-US15925.

PR 10-JUN-1999; 99US-0138453.

XX (MEDI-) MED IMMUNE INC.

PI Adamou JE, Choi GH;

XX WPI; 2001-112197/12.

DR N-ESDB; AAC84742.

XX New vaccines comprising Spi28 or Spi30 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by  
 PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
 XX blood infections  
 XX  
 PS Claim 8; Page 51-54; 54pp; English.  
 XX  
 CC The invention relates to novel immunogenic polypeptides, Sp128 and Sp130  
 CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for  
 CC the treatment and prevention of pneumococcal infections, particularly  
 CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
 CC bronchial, lung or blood infections. The antigens are used as immunogenic  
 CC agents to stimulate an immune response. The antisera and antibodies may  
 CC also be used in diagnosing and treating pneumococcal infections.  
 CC Recombinant polypeptides serve as a mechanism for stimulating production  
 CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
 CC as reagents in other processes such as affinity chromatography. The  
 CC present sequence represents the S. pneumoniae Sp130 polypeptide.  
 XX  
 XX Sequence 773 AA;

Query Match 100.0%; Score 799; DB 22; Length 773;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-71;  
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWISGFEK 60  
 DB 620 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWISGFEK 679  
 QY 61 DAGYVNLKSDTKDVTATVLDKNNISSKSTNNPNK 154  
 DB 680 DAGYVNLKSDTKDVTATVLDKNNISSKSTNNPNK 739  
 QY 121 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
 DB 740 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 773

RESULT 2  
 AAY81710  
 ID AAY81710 standard; Protein; 2120 AA.  
 XX  
 AC AAY81710;  
 XX  
 DT 02-JUN-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae protein sequence ID3.  
 XX  
 KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
 KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
 KW pneumococcal septicaemia; sinusitis; meningitis; therapy.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO200006738-A2.  
 XX  
 PD 10-FEB-2000.  
 XX  
 PF 27-JUL-1999; 99WO-GB02452.  
 XX  
 PR 27-JUL-1998; 98GB-0016336.  
 PR 19-MAR-1999; 99US-0125329.  
 XX  
 PA (MICR-) MICROBIAL TECHNIQS LTD.  
 XX  
 PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;  
 XX  
 DR WPI; 2000-195301/17.  
 DR N-PSDB; AA291806.  
 XX  
 PT Streptococcal proteins and polynucleotides useful for diagnosis,  
 PT treatment and prophylaxis of bacterial infections  
 XX

PS Claim 2; Page 41-42; 76pp; English.  
 XX  
 CC This sequence represents a Streptococcus pneumoniae protein of the  
 CC invention. The proteins (or their homologues, derivatives and/or  
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
 CC compositions comprising the proteins are useful as vaccines and also in  
 CC diagnostic assays. The sequences are useful for the detection or  
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested  
 CC with them. Agents capable of antagonising, inhibiting or interfering with  
 CC the function or expression of the protein or polypeptide are useful in  
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae  
 CC infection. As the sequences can be used to treat S. pneumoniae infection,  
 CC they can be used to treat bacterial pneumonia, which has high rates in  
 CC young children, the elderly, and in patients with predisposing conditions  
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
 CC or with immunosuppressive disorders, especially AIDS. They can also be  
 CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and  
 CC meningitis.  
 XX  
 XX Sequence 2120 AA;

Query Match 100.0%; Score 799; DB 21; Length 2120;  
 Best Local Similarity 100.0%; Pred. No. 9.3e-71;  
 Matches 154; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWISGFEK 60  
 DB 1933 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWISGFEK 1992  
 QY 61 DAGYVNLKSDTKDVTATVLDKNNISSKSTNNPNK 154  
 DB 1993 DAGYVNLKSDTKDVTATVLDKNNISSKSTNNPNK 2086  
 QY 121 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 154  
 DB 2053 EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 2086

RESULT 3  
 ABU01020  
 ID ABU01020 standard; Protein; 2140 AA.  
 XX  
 AC ABU01020;  
 XX  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain protein from coding region #590.  
 XX  
 KW Bacterial meningitis; pneumonia; sepsis; otitis media;  
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KW auditory; respiratory; gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae type 4 strain.  
 XX  
 PN WO200277021-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-IB02163.  
 XX  
 PR 27-MAR-2001; 2001GB-0007658.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Masignani V, Tettelin H, Fraser C;  
 XX  
 DR WPI; 2003-040579/03.  
 DR N-PSDB; ABX06302.  
 XX  
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
 PT



PA (BARA/) BARASH S C.  
 PA (DILL/) DILLON P J.  
 PA (DOUG/) DOUGHERTY B.  
 PA (FANN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Choi GH, Kunech CA, Barash SC, Dillion PJ, Dougherty B, Fannon MR;  
 PI Rosen CA;  
 XX  
 DR WPI; 2002-479261/51.  
 DR N-PSDB; ABQ84825.  
 XX  
 PT New Streptococcus pneumoniae antigens, useful for detecting  
 PT Streptococcus and for preventing or attenuating disease caused by  
 PT Streptococcus infection -  
 XX  
 PS Claim 11; Page 29; 70pp; English.  
 XX  
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
 CC The S. pneumoniae antigens have antibacterial activity and can be  
 CC used in vaccines. The S. pneumoniae antigens can also be used to  
 CC prevent or attenuate a Streptococcal infection in an animal. The  
 CC polynucleotides encoding the S. pneumoniae antigens can be used to  
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
 CC which are used in an example from the present invention.  
 XX  
 SQ Sequence 117 AA;  
 Query Match 77.0%; Score 615; DB 23; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-54;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 38 YKGELEKGYQFDGWISGPEGKDGAGVINLSKDTFIKVPFKIEBKKEENKPTFDVSK 97  
 DB 1 YKGELEKGYQFDGWISGPEGKDGAGVINLSKDTFIKVPFKIEBKKEENKPTFDVSK 60  
 QY 98 KKDNPQVNSQLNESHKEDLQREHSQKSDSTKVATVLDKNNISSKSTTNNPNK 154  
 DB 61 KKDNPQVNSQLNESHKEDLQREHSQKSDSTKVATVLDKNNISSKSTTNNPNK 117  
 RESULT 6  
 AAG81779  
 ID AAG81779 standard; Protein; 746 AA.  
 XX  
 AC AAG81779;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:652.  
 XX  
 DE Staphylococcus epidermidis SR1 strain; infection; diagnosis;  
 KW vaccination; endocarditis.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN WO200134809-A2.  
 XX  
 PD 17-MAY-2001.  
 XX  
 XX 09-NOV-2000; 2000WO-US30782.  
 XX  
 XX 09-NOV-1999; 99US-0164258.  
 XX  
 XX (GLAX) GLAXO GROUP LTD.  
 PA  
 XX Kimmerly WJ;  
 PI  
 XX WPI; 2001-316495/33.  
 DR N-PSDB; AAG82629.  
 XX

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 PT useful for vaccinating against infections, e.g. endocarditis -  
 XX  
 PS Claim 18; Page 208; 218pp; English.  
 XX  
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX  
 SQ Sequence 746 AA;  
 Query Match 14.8%; Score 118; DB 22; Length 746;  
 Best Local Similarity 27.2%; Pred. No. 0.0063;  
 Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps 9;  
 QY 1 DTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYK-----GELEKGYQFDGW--- 51  
 DB 586 DSVNAQSLKP----ITIGNKQIKQSVKSGTKVLPKSHKVMMLTDGELTMP-DMTGWTKE 640  
 QY 52 EISGFE-----GKDGAVIN--LSKDTFIKVPFKIEBKKEENKPTFDVS-----KK 98  
 DB 641 DVLAPEFDTKLKVTSGKNGFVNTQSIKQGIK-----NKKDKIEVLSAEDT 687  
 QY 99 KDNPNQVNSQLNESHKEDLQREHSQKSDSTKVATVLDKNNISSKSTTNN 151  
 DB 688 DDDQETDESDSNKSKKDADEHSNTSSSTKN-----DKSNADSKNDSDD 734  
 RESULT 7  
 ABP39023  
 ID ABP39023 standard; Protein; 778 AA.  
 XX  
 AC ABP39023;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3868.  
 XX  
 DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.  
 XX  
 XX 13-AUG-1998; 98US-0134001.  
 XX  
 PR 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 XX Doucette-Stamm LA, Bush D;  
 XX

```

DR WPI; 2002-381255/41.
DR N-PSDB; ABN91566.
XX
XX
PT Polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX
XX Disclosure; SEQ ID 3868; 267pp; English.
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
XX
SQ Sequence 778 AA;
Query Match 14.8%; Score 118; DB 23; Length 778;
Best Local Similarity 27.2%; Pred.No. 0.0067;
Matches 47; Conservative 23; Mismatches 57; Indels 46; Gaps
QY 1 DTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYK-----GELEKGYQFDGW--- 51
DB 618 DSVNAQSLKP-----IITGNGKQIKQSQSVKSGTKVLPKSKVLMWTGELTWP-DMTGWTKE 71
QY 52 EISGFE-----GKXDAQGVIN--LSKDTFIKVPKPIEKEEENKPTFDVS---KK 98
DB 673 DVLAFEDLTIKIVSTKXGNGFVNTQGISKGQIIK-----NKKIEVLSAEDT 71
QY 99 KDNPNVHNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTN 151
DB 720 DDPOEKTDESDSNKSKKOKADEHDSNTSSSTKN-----DKSNADSKNDSDD 766
RESULT 8
AAG47777
ID AAG47777 standard; Protein; 484 AA.
XX
XX AAG47777;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 60255.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.

```

PT Proteins encoded by chromosome 2 of the human malarial parasite,



PT Plasmodium falciparum, useful as antimalarial vaccines and in the  
 XX diagnosis of P.falciparum infection -  
 PS Disclosure; Page 321-322; 577pp; English.  
 XX  
 CC The present invention describes proteins and their fragments (I) encoded  
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.  
 CC Also described are: (i) nucleotide sequences (ii) encoding (i); and (2)  
 CC vaccines against P. falciparum infection comprising (i) or (ii).  
 CC (i) and (ii) are useful for the development of vaccines against  
 CC P. falciparum infection. (i) and polyclonal antisera or a monoclonal  
 CC antibody raised to immunogens comprising the sequences of (i), are  
 CC useful in the detection of infection with P. falciparum. Furthermore,  
 CC (i) especially when they are rifins or secreted or membrane proteins)  
 CC can aid the identification of drugs to treat or prevent P. falciparum  
 CC infection, or they can be used to identify drug resistance in  
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the  
 CC subsequent identification of proteins encoded by it will help to expand  
 CC our understanding of parasite biology, a process hampered by the  
 CC complexity of the parasitic life cycle, and provide new targets for  
 CC vaccine and drug development. Parasite resistance to drugs and mosquito  
 CC resistance to insecticides have led to a resurgence of malaria in many  
 CC parts of the world, and there is a pressing need for vaccines and new  
 CC drugs. AAU0078 to AAA0287 and AAB1814 to AAB1832 represent nucleotide  
 CC and protein sequences given in the present invention, but which are not  
 CC specifically mentioned within the specification.  
 XX  
 SQ Sequence 665 AA;  
 Query Match 13.1%; Score 105; DB 21; Length 665;  
 Best Local Similarity 24.4%; Pred. No. 0.11;  
 Matches 42; Conservative 34; Mismatches 40; Indels 56; Gaps 10;  
 Qy 1 DTGEVSELKPHRVT-VTTONKEMSSSTIVSEEDFILPVYKGELEKGYQFDGWEI--SGFE 57  
 Db 128 EKNKINKSDLRQNELNLSQSGK-----NEQDI-----NKNEKGKQ-----DISNSNAE 170  
 Qy 58 GKQDAGYVINLSKDTFKPVFKKIEKKE-----EENKPTFD-----VSKKKDNQP 103  
 Db 171 NKKD-----VKEGVKELEKKEKKEKISDDHKVEENKSDDDHKVEENKSDDHK 218  
 Qy 104 VNHSQLNESHKREDLQR-EHHSQKSDSTKDTATVTLDRKNISSKSTTNNPNK 154  
 Db 219 VEENKSDDHKIEVKVVEEHEDEE-----DKKEKSEKNKNDENK 261  
 RESULT 10  
 ABB61977  
 ID ABB61977 standard; Protein; 564 AA.  
 AC ABB61977;  
 XX  
 XX 26-MAR-2002 (first entry)  
 DT Drosophila melanogaster polypeptide SEQ ID NO 12723.  
 DE Drosophila melanogaster polypeptide; cell signalling; insecticide;  
 KW pharmacological.  
 XX  
 XX Drosophila melanogaster.  
 OS WO200171042-A2.  
 XX  
 XX 27-SEP-2001.  
 PD  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI

XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL06080.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 XX interactions -  
 PS Disclosure; SEQ ID NO 12723; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-ABL30511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins  
 CC (AB157737-AB172072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 564 AA;  
 Query Match 12.7%; Score 101.5; DB 22; Length 564;  
 Best Local Similarity 24.5%; Pred. No. 0.19;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 Qy 31 EDFILPVYKGELEKGYQFDGW-----EISFEGKKDAGYVI-----NLSKDTIK 75  
 Db 78 EDLDTPLSESFRFSK--VPDGVVDEHDEHGDHGVQEPSGEALDDHDEHDDHDEDEE 135  
 Qy 76 PVFKKIEEKEENKPT-----PDVSKKKDNPOVNHSQLNESHKREDLQREHHSQKSDS 129  
 Db 136 PLTEEEEEEETPEDEPADEVEEDEDENNA--GENITADAEDEEEDND 193  
 Qy 130 TKDVTATVLDKNISSKST 148  
 Db 194 EGTVEATVEATTEATTEAT 212  
 RESULT 11  
 AAE20967  
 ID AAE20967 standard; Protein; 2060 AA.  
 XX  
 AC AAE20967;  
 XX  
 XX 01-JUL-2002 (first entry)  
 DT Staphylococcus lugdunensis von Willebrand factor binding protein.  
 DE Von Willebrand factor binding protein; vWb; immunogen; antibacterial;  
 KW vaccine; infection.  
 XX  
 XX Staphylococcus lugdunensis.  
 OS Key Location/Qualifiers  
 FH Peptide 1..47  
 FT /label= Signal\_peptide  
 FT Protein 48..2060  
 FT /note= "Mature von Willebrand factor binding protein"  
 XX  
 XX WO200228892-A1.  
 XX  
 XX 11-APR-2002.  
 PD  
 XX 06-APR-2001; 2001WO-SE00766.  
 PF  
 XX 04-OCT-2000; 2000SE-0003573.  
 PR  
 XX (BIOS-) BIOTAPRO AB.  
 PA  
 XX Guss B, Frykberg L, Jacobsson K, Ahlen J, Nilsson M;  
 PI

```
XX WPI; 2002-304928/34.
DR N-PSDB; AADJ3371.
XX
PT New von Willebrand factor binding protein from Staphylococci, useful
PT for determining and treating staphylococcal infection
XX
XX Claim 3; Page 35-41; 53pp; English.
XX
CC The present invention relates to von Willebrand factor binding protein or
CC polypeptide (vwb) from Staphylococci. The vwb and immunogens of vwb are
CC useful in vaccines to combat infections caused by Staphylococci. The
CC invention is also useful for detection of staphylococcal infection and
CC purifying von Willebrand factor from a complex solution. The present
CC sequence is Staphylococcus lugdunensis vwb1 protein.
XX
SQ Sequence 2060 AA;
Query Match 12.1%; Score 96.5; DB 23; Length 2060;
Best Local Similarity 25.7%; Pred. No. 3.8;
Matches 39; Conservative 25; Mismatches 49; Indels 39; Gaps 9;
QY 4 EYSELKPHRVTVTIQ-NGKEMSTIVSEEDFILPVYKGELEKGYPDG--WBISGFEGKK 60
DB 1914 DIELK---ITIVDTNGRE-----IVPSRKGLPP-EQFIGODWQYTGHK--- 1955
QY 61 DAGYVNLKSTQFIKPVFKKIEE-----KKEENKPTFDVSKKQNPQVNHSQLNESH 113
DB 1956 -----IEKDGITTVYKKVENAVPAKQLKTKHN--TQSBQFKHTPQVKQLVKVHN 2006
QY 114 RKE--DLQREHSQKSDSTKDTATVLDKNNI 143
DB 2007 VKQRSIEKSEHTDMHVSLEPGETANKGL 2038
XX
RESULT 12
AAB18272
ID AAB18272 standard; Protein; 2500 AA.
XX
AC AAB18272;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:129.
XX
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
XX
PA (CARU/) CARUCCI D.
XX
PA (GARD/) GARDNER M.
XX
PA (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection
XX
XX Disclosure; Page 302-309; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 2500 AA;
Query Match 12.0%; Score 96; DB 21; Length 2500;
Best Local Similarity 26.1%; Pred. No. 5.5;
Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;
QY 9 KPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYOFGWEISGFEGKDGAVINL 68
DB 2173 KPYKIT---ENKK-----NEGNEILKKYSTENEKKNYDKEQNECILDKDTQCNVNT 2223
QY 69 -----SKDTFIKPVFKKIEEENKPTFDVSKKQNPQVNHSQLNE----SHRKED 117
DB 2224 KEKNLNDNKKSPFSNKKVLEEEKSDDKRD---DKNDNTRKNNLNDNKKSPFSNKKV 2280
QY 118 LQREHSQKSDSTKDTATVLDKNNISSKSTTNNPK 154
DB 2281 LEESE---KSDDKRD-----DKNDNTRKNNLNDNKKSPFSNKKV 2308
XX
RESULT 13
AAG37134
ID AAG37134 standard; Protein; 408 AA.
XX
AC AAG37134;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45610.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
```

PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 25-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132485.  
PR 05-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.

```
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.7%; Score 93.5; DB 21; Length 408;
Best Local Similarity 25.0%; Pred. No. 0.79;
Matches 39; Conservative 28; Mismatches 58; Indels 31; Gaps 8;

QY 4 EYSELKP-HRVVTIQNGKEMSTIVSEDFILPVYKGEKGYQFDGWEISFGKKDA 62
DB 201 EKDPLKPHVSAFLVYANERRAAURENKSVVEVAK-----ITGEWNKNSDKKA 252

QY 63 GY--VINLSKDTFIKVPFKKIBKKKEENKPTFDVSKKNDPNQVNHSQLNESHKEDLQR 120
DB 253 PYEKVAKKNKEIYQ-AMEYKTRKEE-----ALSOKE-----EEELLKLHKQEAQM 301

QY 121 EHSQKSDSTKQVATVLNKNKSNISSTNN--PNK 154
DB 302 LKKKERTDN-----LIKKEATKKKKNNVDPNK 330

RESULT 14
AAG371133
ID AAG371133 standard; Protein; 456 AA.
XX
AC AAG371133;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45609.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
```

```
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.7%; Score 93.5; DB 21; Length 456;
Best Local Similarity 25.0%; Pred. No. 0.92;
Matches 39; Conservative 28; Mismatches 58; Indels 31; Gaps 8;

Qy 4 EVSEKPD-HRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGWEISGPEGKDA 62
Db 249 EKDPLKPKHPVSAFLVTVANERRAALREENKSVVEVAK-----ITGEWKNLSDKKKA 300

Qy 63 GY--VINLSKDTFKIPVKKIEEKEENKPTFDVSKKKNQPNVHSQLNESHKEDLQR 120
Db 301 PYEKVAKKNKETYLIQ-AMEEYKRTKEE-----ALSQKE-----EBELLKHQEQALQM 349

Qy 121 EHSOKSDSKDVTATVLDKNNISSKSTTN--PNK 154
Db 350 LKKKERTDN-----LIKKEKATYKKKNENVDNPK 378

RESULT 15
AAG37132
ID AAG37132 standard; Protein; 476 AA.
XX
AC AAG37132;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 45608.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
FN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
```

PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140895.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142855.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.

```
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.7%; Score 93.5; DB 21; Length 476;
Best Local Similarity 25.0%; Pred. No. 0.98; Mismatches 58; Indels 31; Gaps 8;
Matches 39; Conservative 28;

QY 4 EVSELPK-HRVTVTIQNGKEMSTTVSEEDFILPVVKGLEKGYQFDGWEISGFGKKDA 62
DB 269 EKDPLKPHVSAFLVYANERAAALREENKSVVEAK-----ITGEWKNLSDDKKA 320
QY 63 GY--VINLSKDTFIKPVFKIEEKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR 120
DB 321 PYEKVAKKNKETYLO-AMBEYKRTKEE-----ALSQKE-----EEELLKLHKQEAQM 369
QY 121 EHSQKSDSTKDVATVLDKNNISKSSTNN--PNK 154
DB 370 LKKEKTDN-----LIKKEKATKKKNENVDPNK 398

RESULT 16
ABG16636
ID ABG16636 standard; Protein; 2519 AA.
XX
AC ABG16636;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #16627.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
WPI; 2001-639362/73.
XX
```

```
DR N-PSDB; AAS80823.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 46995; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2519 AA;

Query Match 11.6%; Score 92.5; DB 22; Length 2519;
Best Local Similarity 31.0%; Pred. No. 13;
Matches 22; Conservative 19; Mismatches 17; Indels 13; Gaps 3;

QY 68 LSKDTFIKPVFKIEEKEENKPTFDVSKKDNPNQVNHSQLNESHKEDLQR-----E 122
DB 689 VKKETKVP-----EDKKEEKEPKKEVAKEDKTPI---KKEEKPKEVKEVKEIK 740
QY 123 HSQKSDSTKDV 133
DB 741 KEEKKEPKKEV 751

RESULT 17
AAB42897
ID AAB42897 standard; Protein; 209 AA.
XX
AC AAB42897;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF2661 polypeptide sequence SEQ ID NO:5322.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
```



PN WO200058473-A2.  
XX 05-OCT-2000.  
XX 31-MAR-2000; 2000WO-US08621.  
XX 31-MAR-1999; 99US-0127607.  
PR 02-APR-1999; 99US-0127636.  
PR 05-APR-1999; 99US-0127728.  
PR 30-MAR-2000; 2000US-0540763.  
XX (CURA-) CURAGEN CORP.  
XX Shimkets RA, Leach M;  
XX WPI; 2000-602362/57.  
XX N-PSDB; AAC77106.  
XX Novel nucleic acids and peptides derived from open reading frame X,  
PT useful for treating e.g. cancers, proliferative disorders,  
PT neurodegenerative disorders and cardiovascular disease -  
XX Claim 11; Page 4497-4498; 5507pp; English.  
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;  
CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;  
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;  
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotropic;  
CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;  
CC antithyroid; and antianemic. The sequences can be used for determining  
CC the presence of or predisposition to, or preventing or treating  
CC pathological conditions associated with an ORFX-associated disorder. The  
CC nucleic acids can be used to express ORFX proteins in gene therapy  
CC vectors. The proteins and nucleic acids may be used to treat cancers,  
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral  
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance  
CC coagulation; to inhibit thrombosis; and as a contraceptive.  
XX Sequence 209 AA;  
SQ  
Query Match 11.3%; Score 90.5; DB 21; Length 209;  
Best Local Similarity 28.1%; Pred. No. 0.62;  
Matches 47; Conservative 17; Mismatches 56; Indels 47; Gaps 9;  
QY 3 GEVSELKPHRVTVTQNGKEMSTIVSE--EDFILPVYKGELEKGYQF-----DGWE--- 52  
Db 18 GELQLUGSQL-----HGKSDSNVYTEKKEITALLRERUTELERKUTFTQQSDLMERLY 72  
QY 53 -----TSGFEGKKDAG-----YVINLSKDTFIPVKFKIBKKEENKPTFDVSKKD 100  
Db 73 VEAKDQNGKQGTGDKKKGGGSHRAKNKSKFVLSV-----KEITFDMKNST 120  
QY 101 NPQVNH-----SQLNESHKEDLQREHSQKSD--STKDVATVLDK 140  
Db 121 KEFVRHHEKIKQAKEA-VKENLKFPDSVSKSTFRHFKDTTKNIFDE 166  
RESULT 18  
AAY44364  
ID AAY44364 standard; protein; 757 AA.  
XX AAY44364;  
AC AAY44364;  
XX 14-MAR-2000 (first entry)  
DT  
XX

DE Human cell cycle regulation protein-5.  
XX CECRP-5; cell cycle regulation protein-5; cell proliferation;  
KW cell proliferative disease; cancer; atherosclerosis; cirrhosis;  
KW hepatitis; psoriasis; immune system disorder; allergy; asthma;  
KW acquired immune deficiency syndrome; Crohn's disease; Blast method;  
KW rheumatoid arthritis; gene therapy; chromosomal mapping.  
OS Homo sapiens.  
XX Key Location/Qualifiers  
FH Modified-site 4  
FT Modified-site /note= "Potential glycosylation site"  
FT Modified-site 63  
FT Modified-site /note= "Potential glycosylation site"  
FT Modified-site 479  
FT Modified-site /note= "Potential glycosylation site"  
FT Modified-site 498  
FT Modified-site /note= "Potential glycosylation site"  
FT Modified-site 31  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 43  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 81  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 96  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 99  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 130  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 158  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 162  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 163  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 184  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 186  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 188  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 271  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 277  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 287  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 289  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 323  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 343  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 367  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 370  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 432  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 488  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 499  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 207  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 277  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 298  
FT Modified-site /note= "Potential phosphorylation site"  
FT Modified-site 370  
FT Modified-site /note= "Potential phosphorylation site"

FT	Modified-site	419	/note=	"Potential phosphorylation site"
FT	Modified-site	474	/note=	"Potential phosphorylation site"
FT	Modified-site	488	/note=	"Potential phosphorylation site"
FT	Modified-site	499	/note=	"Potential phosphorylation site"
FT	Modified-site	528	/note=	"Potential phosphorylation site"
FT	Modified-site	532	/note=	"Potential phosphorylation site"
FT	Modified-site	539	/note=	"Potential phosphorylation site"
FT	Modified-site	609	/note=	"Potential phosphorylation site"
FT	Modified-site	624	/note=	"Potential phosphorylation site"
FT	Modified-site	730	/note=	"Potential phosphorylation site"
FT	Modified-site		/note=	"Potential phosphorylation site"

XX PN WO9964593-A2

XX  
PD 16-DEC-1999.

08-JUN-1999: 99WQO-US12906-XX PF

XX  
PR 08-JUN-1998: 98IIS-0088695

XX PA (TNCY-) TNCYTE PHARM TNC-

XX Bandman O, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR;  
PI Patterson C;  
PI

WPI: 2000-105887/09.

DR N-PSDB; AAZ29484.

Novel regulatory proteins, for diagnosis, treatment and prevention of cell proliferative and immune system diseases -

PS Claim 1; Page 70-71; 87pp; English.

The present sequence is cell cycle regulation protein-5 (CECRP-5). Blast analytical method was used to identify this protein. CECRPs are activators of cell proliferation or inhibitors of cellular processes that modulate proliferation. They are used to treat or prevent cell proliferative diseases like cancers, atherosclerosis, cirrhosis, hepatitis, psoriasis, immune system disorders (e.g. acquired immune deficiency syndrome, allergy, asthma, Crohn's disease, rheumatoid arthritis). Antibodies are raised to screen for specific binding agents. The corresponding nucleic acid is used in gene therapy, chromosomal mapping and isolation of related sequences.

XX	Sequence	757 AA:
SO		

Query Match	11.3%	Score 90.5;	DB 21;	Length 757;
Best Local Similarity	28.1%	Pred. No. 3.7;		
Matches	47: Conservative	17: Mismatches	56: Indels	47: Deletions

**Qy** 3 GEVSELKPRVTVTIQNGKMSSTIVSE--EDFILPVYKEGLEKGQF-----DGWE--- 52  
||:|:||:||||:||||:||||:

**Dp** 398 GELOOLSGSOL-----HGKSDSPNYVTEKKEIAILPERLLELERKLTFEORSDLWRLY 452

Qy	53	-----	ISGFECKDAG	----	YVINLSKDTFIKVPFKIEEKKKEENKPTFDVSKKD	100
		:		:		
Db	453	VEAKDNGKOGTDGKKKGGRGSHRAKNKSKETFLGSV	-----	RETFDAMKNST	500	

101 NPQVNH-----SOLNESHKEDLOREEHSOKSD--STKQVTAIVLTK 140

db 501 KEFVRHHEKTKOAKEA - VKENLKKFSDSVKSTFRHFKOTTKNIFE 546

## RESULT 19

AAB53319

ID AAB53319 standard; Protein; 758 AA.

AC AAB53319:

09-MAR-2001 (first entry)

DE Human colon cancer antigen protein sequence SEQ ID NO:859.

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW identification; cytostatic; cardiotoxic; neuroprotective; vulnery;  
KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
KW nephrotropic; antineflect; antibacterial; gene therapy; wound;  
KW neural disorder; immune system disorder; muscular disorder;  
KW reproductive disorder; gastrointestinal disorder; renal disorder;  
KW infectious disease; cardiovascular disorder.

XX Homo sapiens.

XX  
PN  
W0200055351-A1XX  
PD  
21-SEP-2000XX  
PF 08-MAR-2000. 2000WC-IIS05883XX  
12-MAP-1999. 99JIS-0124270  
PPXX  
(HTMA) HUMAN GENOME SCI TNC  
PA

XX	Docos	CA	Pubos	CM.
NT				

[illegible]

DR N-PSDB; AAC98076.

PT Colon cancer associated gene sequences, referred to as colon cancer  
PT antigens, useful for the treatment, prevention, and diagnosis of colon  
PT disorders such as colon cancer -

XX  
PS  
Claim 11: Page 1408-1411: 2104pp: English.

AAC97991 to AAC98763 encode the human colon cancer associated proteins, called human colon cancer antigens, given in AAB53234 to AAB54006. The human colon cancer antigens can have cytostatic, cardioactive, muscular; neuroprotective, immunomodulatory, gynaecological, gastrointestinal, and vulvular, nephrotropic, antineoplastic and antibacterial activities, and can be used in gene therapy. The colon cancer antigen polynucleotides, proteins and antibodies to the proteins are useful for the prevention, treatment and diagnosis of colon disorders, such as colon cancer. The polynucleotides may be used in diagnostics and research, such as for chromosome identification, and as hybridisation probes. The proteins may also be used to prevent diseases such as neural disorders, immune system disorders, muscular disorders, reproductive disorders, gastrointestinal disorders, wounds, renal disorders, infectious diseases, and cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent sequences used in the exemplification of the present invention.

XX  
S0  
Sequence 758 AA:

Query Match	11.3%	Score 90.5;	DB 21;	Length 758;
Best Local Similarity	28.1%	Pred. No. 3.8;		
Matches	47: Conservative	17: Mismatches	56: Indels	47: Indels

Qy 3 GEVSELPKPRVTVTIQNGKMSSTIVSE--EDFILPYKGELEKGYQF-----DGWE--- 52

Dp 399 GELOOLSGSOL-----HGKSDSPNVTEKETAIRBELTELRKLTFGOORSDLWERLY 45

Qy	53	-----ISGFGEKKDAG-----YVINLSKDTFIKVPFKKIEEKKEEENRPTTFDVSKKKD	100
Dp	454	VEAKDNGKGGTGGKKKGGGSGSHRAKNNKSKETFLGSLV-----KETFDAMKNST	501

```
QY      101 NPQVNH-----SQLNESHKEDLQREEHSQKSD--STKDVTATVLDK 140
```



```

Db 396 MVEGQVVRTISKDAKNNTRTIIPYVEGKTLDAIVKVHVKTIDYDGOYHVRIVDKAFT 455
Qy 75 KPVFKKIEEKEENKPTFDV-----SKKDNPOVNHSQLNESHKEDLQ-----REHSQ 125
Db 456 KANTDKSNKQSQDQNSAKKEATPATPSKPTSPVEKESQKQSDKNKQLPSVEKENDA 515
Qy 126 KSDSTKQVDT-ATVLDKNNISSKSTNNPNK 154
Db 516 SSESCKDTPATKPTKGEVSSSTT--PTK 543

```

## RESULT 22

ABP56879  
ID ABP56879 standard; Protein; 654 AA.

XX AC ABP56879;

XX DT 07-APR-2003 (first entry)

XX DE Staphylococcus epidermidis KrKn protein SEQ ID NO:10.

XX KW Staphylococcus aureus; Staphylococcus epidermidis; MSCRAMM; antibody;  
microbial surface component recognising adhesive matrix molecule;  
surface protein; infection; antibacterial; antiinflammatory; vaccine;  
immunosuppressive; antiarthritic; gene therapy; pneumonia; endocarditis;  
septic arthritis; biomaterial related infection.

XX OS Staphylococcus epidermidis.

XX PN WO2002102829-A2.

XX PD 27-DEC-2002.

XX PF 17-JUN-2002; 2002WO-US19220.

XX PR 15-JUN-2001; 2001US-298098P.

XX PA (INHI-) INHIBITEX INC.

XX PA (QUEE-) QUEEN ELIZABETH COLLEGE DUBLIN.

XX PA (UTPA-) UNIV PAVIA.

XX PI Foster TJ, Roche F, Patti JM, Hutchins JT, Hall A, Domanski P;

XX PI Patel P, Syribeys P, Speziale P;

XX XX WPI; 2003-167481/16.

XX DR N-PSDB; ABZ22903.

XX PT New isolated antibody that binds to a staphylococcal surface protein,  
useful for treating or preventing Staphylococcus aureus infections,  
such as pneumonia, septic arthritis, endocarditis or biomaterial  
related infections -

XX PS Claim 1; Page 29-30; 122pp; English.

XX CC The present invention describes an isolated antibody (I) that binds to a  
staphylococcal surface protein selected from any of the 12 sequences of  
354-2283 amino acids given in ABP56875 to ABP56886. Also described:  
CC (1) an isolated antisera (II) comprising (I); (2) a diagnostic kit (III)  
comprising (I) and means for detecting binding by (I); (3) diagnosing  
CC (MI) an infection of Staphylococcus aureus by adding (I) to a sample  
suspected of being infected with the infection, and determining if  
CC antibodies have bound to the sample; (4) a pharmaceutical composition  
CC (IV) for treating or preventing an infection of S. aureus comprising (I),  
and a vehicle, carrier or excipient; (5) treating (M2) or preventing an  
CC infection of S. aureus by administering (I) to a human or animal patient;  
CC (6) producing (M3) an immunological response by administering to a human  
or animal an immunogenic amount of the isolated surface protein; (7) an  
CC isolated active fragment (V) from the A domain of the DsaA protein; and  
CC (8) a vaccine (VI) for treating or preventing an infection of S. aureus  
comprising the surface protein in an amount effective to elicit an immune  
CC response, and a vehicle, carrier or excipient. (I) has antiinflammatory,  
CC antibacterial, immunosuppressive and antiarthritic activities, and can be

CC used in gene therapy. The antibody, composition and vaccine are useful  
CC for treating or preventing Staphylococcus aureus infection in a human or  
CC animal, such as pneumonia, septic arthritis, endocarditis or biomaterial  
CC related infections. The present sequence represents Staphylococcus  
CC epidermidis KrKn protein, which is used in the exemplification of  
CC the present invention.

XX SQ Sequence 654 AA;

Query Match 11.3%; Score 90; DB 24; Length 654;  
Best Local Similarity 21.4%; Pred. No. 3.4;  
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;

```

Qy 4 EVSELKPHRTV--TIQNGKEMSTIVSEEDFILPVYKGELE-KGYQF-----DGWEIS 54
Db 350 KMTDLQDTKYVYVESVENNESMDTFFVKH-----PIKTGMLNGKKYVMYMETTNDYWKDF 404
Qy 55 GFEGKK-----DAGYVNL-SKOTFI 74
Db 405 MVEGQVVRTISKDAKNNTRTIIPYVEGKTLDAIVKVHVKTIDYDGOYHVRIVDKAFT 464
Qy 75 KPVFKKIEEKEENKPTFDV-----SKKDNPOVNHSQLNESHKEDLQ-----REHSQ 125
Db 465 KANTDKSNKQSQDQNSAKKEATPATPSKPTSPVEKESQKQSDKNKQLPSVEKENDA 524
Qy 126 KSDSTKQVDT-ATVLDKNNISSKSTNNPNK 154
Db 525 SSESCKDTPATKPTKGEVSSSTT--PTK 552

```

## RESULT 23

ABBS8769  
ID ABBS8769 standard; Protein; 1183 AA.

XX AC ABBS8769;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 3099.

XX KW Drosophila; developmental biology; cell signalling; insecticide;  
pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL02872.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more  
genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -

XX PS Disclosure; SEQ ID NO 3099; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent  
capable of detecting 1000 or more genes from Drosophila. The invention is  
useful in developmental biology and in elucidating cell signalling and  
cell-cell interactions in higher eukaryotes for the development of  
insecticides, therapeutics and pharmaceutical drugs. The invention  
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AB57737-AB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1183 AA;  
 Query Match 11.3%; Score 90; DB 22; Length 1183;  
 Best Local Similarity 24.7%; Pred. No. 7.8;  
 Matches 36; Conservative 24; Mismatches 54; Indels 32; Gaps 5;  
 QY 9 KPHRVTTVQNGKMSSTIVSEDFILPVYKGELEKGYQDFGWEISGFGKDGAVINL 68  
 DB 310 KPEVKTAKDNKDK---PLAEDSKPEVPKETS-----TGGKKGGA----- 352  
 QY 69 SKDTFKVPFKIEEKKEENKPTFDVSKKDNQVNHSQLNESHKDLQREHSQ--- 125  
 DB 353 -----KPAEK--SKEEKEDSTTSKKEKADSPANNKEQIQGVKKTSFEDASNSHK 404  
 QY 126 ----KSDSTKDVATVLDKNNISSKS 147  
 DB 405 SVTAKKSKKDEASTQVKSDESPEA 430  
 RESULT 24  
 ID AAY03190 standard; Protein; 402 AA.  
 AC AAY03190;  
 XX  
 DT 16-JUN-1999 (first entry)  
 XX  
 DE S. aureus trigger factor protein sequence.  
 XX  
 KW Trigger factor; tig; diagnosis; immune response; bacterial infection;  
 KW S. aureus infection; otitis media; toxic shock syndrome; conjunctivitis;  
 KW wound infection; impetigo; septic arthritis; therapy.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 PN EP905233-A2.  
 XX  
 PD 31-MAR-1999.  
 XX  
 PF 21-AUG-1998; 98BP-0306697.  
 XX  
 PR 29-DEC-1997; 97US-0999339.  
 PR 04-SEP-1997; 97US-0057511.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX  
 PI Burnham MKR, Fosberry A, Hodgson JE, Jaworski DD;  
 PI Lawlor EJ, Rosenberg M, Traini CM, Wang M, Ward J;  
 PI Warren RL;  
 XX  
 DR WPI; 1999-192663/17.  
 DR N-PSDB; AAX28175.  
 XX  
 PT New Staphylococcus aureus trigger factor (tig) polypeptide and  
 PT polynucleotide, useful as diagnostic reagents and for prevention and  
 PT treatment of Staphylococcus aureus infections  
 XX  
 PS Claim 23; Page 8; 31pp; English.  
 XX  
 CC This sequence is the Staphylococcus aureus trigger factor (tig)  
 CC polypeptide of the invention. Tig polypeptides and polynucleotides are  
 CC useful for diagnosing diseases related to over or underexpression of tig  
 CC protein by identifying mutations in the tig gene, or determining tig  
 CC polypeptide or mRNA expression levels due to an infection of an organism  
 CC with the tig gene. They can diagnose the stage and type of infection. Tig  
 CC polypeptides are also useful for screening compounds which affect

CC activity of the protein by measuring the binding to tig and observing the  
 CC stimulation or inhibition of the polypeptide function. These can be used  
 CC in treatment to inhibit or enhance tig activity. In addition to direct  
 CC administration of tig polypeptides to treat conditions associated with a  
 CC lack of tig polypeptide, or direct administration of antisense sequences  
 CC to prevent expression. Tig polypeptides (administered directly, in a  
 CC vector and as a vaccine) and antibodies induce an immune response to  
 CC immunise and prevent disease. Diseases diagnosed, prevented or treated  
 CC include: bacterial infections, especially Staphylococcus aureus  
 CC infections which cause otitis media, toxic shock syndrome,  
 CC conjunctivitis, wound infection, impetigo and septic arthritis etc. Tig  
 CC polypeptides, polynucleotides and their (ant)agonists can prevent  
 CC adhesion of bacteria to matrix proteins, and are useful for use on wounds  
 CC and body implants to prevent bacterial infection.  
 XX  
 SQ Sequence 402 AA;  
 Query Match 11.2%; Score 89.5; DB 20; Length 402;  
 Best Local Similarity 22.1%; Pred. No. 2;  
 Matches 50; Conservative 33; Mismatches 60; Indels 83; Gaps 13;  
 QY 1 DTGEVSELKPHRVTVT-IQNGKE--MSSTIVSEDFILPVYKG-ELEK----- 44  
 DB 53 DETDIKPVQAQPEVSVTQIEKGKDFEATVTVEPEVKLGDKGLBIKQETELSDDELQE 112  
 QY 45 -----GYQFDGWEISG--FEGKDGAGVIVNLSKDTFIKP 76  
 DB 113 AIDHSLGHLAEMVVKEDGVVNGDVTNIDFSG-SVDGEFEGGQAGYDLEIGSGSFI-P 170  
 QY 77 VPK-----KIEEKKE-----EE--NKPTFDVS-----KKDNQVNHSQLN- 111  
 DB 171 GFEEQLEGKVDDEKDVVTPPEYHABELAGKATFKTKNEIKFKEVPELTDEIANEL 230  
 QY 112 -----SHRKEDLQREHSQKSDTKDVATVLDKNNISSKTTN 150  
 DB 231 DAENTVDEYKENLKRLEAQKATDAENV-----EKEBAITKATDN 271  
 RESULT 25  
 ID AAY35091 standard; Protein; 511 AA.  
 XX  
 AC AAY35091;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE Chlamydia pneumoniae transmembrane protein sequence.  
 XX  
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO9927105-A2.  
 XX  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-IB01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Griffais R;  
 XX  
 WPI; 1999-357842/30.  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 XX  
 PS Page 975-976; Disclosure; 1912pp; English.  
 XX

WPI; 2001-656860/75.  
N-PSDB; ABL02807.

PT interactions -  
XX PS Disclosure; SEQ ID NO 2904; 2lpp + Sequence Listing; English.  
XX PS  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX PS  
SQ Sequence 1408 AA;  
Query Match 11.2%; Score 89.5; DB 22; Length 1408;  
Best Local Similarity 24.7%; Pred. No. 11;  
Matches 48; Conservative 23; Mismatches 62; Indels 61; Gaps 9;  
QY 19 NGKE-----MSSITVSEDF--ILPVYKGELEKGYQDGEISGFEGK--KDAGYV----- 65  
Db 646 NGKQKLNASKEYTKSEKDFKEIVSVSKE-ERNDQESKELENQDGLPKDQDYLDVKKPP 704  
QY 66 -----INLSKDTFIKPVFK-----KIEKKEEENKPTFDVSKKKNQPNQVNHSLN 110  
Db 705 MESKEYSEPLAVKLGEDVSKSPSNESSTKISGVKGE--PKMNGELPKTGQVQKPEKS 762  
QY 111 ESHRK-EDLQREHSOKSPSTKQDVT-----ATVLDK 140  
Db 763 KTOAKTEFLAKVQSELSKSVKDVTPKEOTTKEETHKGFSPKREATPKSEPKGTTEVDK 822  
QY 141 NNISKSTNNPNK 154  
Db 823 SNDVSVIKSEPDE 836  
RESULT 28  
ID ABJ18979  
XX ABJ18979 standard; Protein; 645 AA.  
AC ABJ18979;  
XX  
DT 06-MAR-2003 (first entry)  
XX  
DE Pathogen specific antigen related staphylococcal protein SEQ ID No 142.  
XX  
KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;  
KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;  
KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
KW autoimmune disease; HIV; hepatitis.  
XX  
OS Staphylococcus sp.  
XX  
FN WO200259148-A2.  
XX  
PD 01-AUG-2002.  
XX  
PF 21-JAN-2002; 2002WO-EP00546.  
XX  
PP 26-JAN-2001; 2001AT-0000130.  
XX  
PR (CIST-) CISTEM BIOTECHNOLOGIES GMBH.  
XX  
PA Meinke A, Nagy E, Von Ahsen U, Klade C, Henics T, Zauner W;  
PI Minn DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;  
PI Tempelmaier B;  
XX  
DR WPI; 2003-075410/07.  
XX  
PT Identifying, isolating and producing hyperimmune serum-reactive

PT antigens from a pathogen, for preparing vaccine or medicament for  
PT treating or preventing e.g. staphylococcal infections, comprises  
PT providing antibody preparation -  
XX  
XX Claim 21; Page 168; 252pp; English.  
XX  
CC The invention relates to a novel method for identifying, isolating and  
CC producing hyperimmune serum-reactive antigens from a pathogen, tumour,  
CC allergen, a tissue or host prone to auto-immunity, where the antigens  
CC are used in a vaccine, comprises providing antibody preparation from a  
CC plasma pool of a type of animal, or individual sera with antibodies  
CC against the specific pathogen, tumour, allergen, tissue or host prone to  
CC auto-immunity. The hyperimmune serum-reactive antigens comprising any of  
CC the 62 sequences of 53-2261 amino acids fully defined in the  
CC specification, or their hyperimmune fragments are useful for the  
CC manufacture of a pharmaceutical preparation, particularly a vaccine  
CC against staphylococcal infections or colonisation against S. aureus or S.  
CC epidermidis. The preparation of antibodies is useful for the manufacture  
CC of a medicament for treating or preventing staphylococcal infections or  
CC colonisation against S. aureus or S. epidermidis. The antibody  
CC preparations may also be used for diagnostic and imaging purposes. Other  
CC conditions that can be treated include cancer, autoimmune diseases or  
CC infections caused by viral (e.g. HIV, hepatitis A, B or C), fungal or  
CC protozoan pathogens. This sequence represents a staphylococcal protein  
CC relating to the method for identifying and producing pathogen specific  
CC antigens of the invention.  
XX  
SQ Sequence 645 AA;  
Query Match 11.1%; Score 89; DB 24; Length 645;  
Best Local Similarity 21.4%; Pred. No. 4.2;  
Matches 45; Conservative 32; Mismatches 67; Indels 66; Gaps 10;  
QY 4 EVSELKPHRVTV--TIQNGKMSITVSEEDFILPVYKGELE-KGYQF-----DGWEIS 54  
Db 341 KMTDLQDTKYVYVESVENNESMMDTFVKH-----PIKTGMLNGKMYMVTNDDYWKDF 395  
QY 55 GFEGKK-----DAGVINL-SKDTFI 74  
Db 396 MVYEQRVRTISKAKNTRTIIFPVYEGKTLVDALVKVHVKTIDYDGHVIRIVDKAFT 455  
QY 75 KPVFKKIEKKEEENKPTFDV-----SKKKNQPNQVNHSLNESHKEDLQ-----REEHSQ 125  
Db 456 KANTDKSNKKEQDNSAKKEATPATPSKPTSPVEKESQKDSQKDDNKQLPSVEKENDA 515  
QY 126 KSDSTKQVTV--ATVLDKNNISSKSTNNPNK 154  
Db 516 SSESCKGVTLATKPTKGEVESSTT--PTK 543  
RESULT 29  
ID ABB64828  
XX ABB64828 standard; Protein; 281 AA.  
AC ABB64828;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 21276.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
FN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.



```
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-658660/75.
XX N-PSDB; ABL08931.
DR New isolated nucleic acid detection reagent for detecting 1000 or more
XX Genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 21276; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 281 AA;
XX
XX Query Match 11.1%; Score 88.5; DB 22; Length 281;
XX Best Local Similarity 22.1%; Pred. No. 1.5;
XX Matches 31; Conservative. 34; Mismatches 52; Indels 23; Gaps 5;
XX
XX Qy 19 NGKSSSTIVSEEDFILPVYKGELEKGYQFDGWEISFGKKGADGYVI-----NLS 69
XX Db 137 NFEFTNMTTQDDLLKNYI-SLLDK-----DNEGAIKSELGWNVRALGRQPNESIA 188
XX
XX Qy 70 KDTFTKPVFKKIEE-KKEEENKPTFDVSKKDNPOVNHSQLNE-----SHRKEDLQREEH 123
XX Db 189 KEFCNVLIRKWHDTNKEELRDARFVFDKENNGYISITELRAVFMALGEXLEDELEEM 248
XX
XX Qy 124 SKSDSTKDVATVLDKNNI 143
XX Db 249 IREYDLQDNHINFESNNM 268
XX
XX RESULT 30
XX AAG85008
XX ID AAG85008 standard; Protein; 1141 AA.
XX
XX AC AAG85008;
XX
XX DT 11-SEP-2001 (first entry)
XX
XX DE Shrimp white spot Bacilliform virus (WSBV) protein 99.
XX
XX KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
XX KW antiviral agent; gene expression; antisense construct;;
XX KW transgenic viral resistant shrimp.
XX
XX OS White spot syndrome virus.
XX
XX PN WO2001138351-A2.
XX
XX PD 31-MAY-2001.
XX
XX PF 08-NOV-2000; 2000WO-US28888.
XX
XX PR 24-NOV-1999; 99CN-0124717.
XX
XX (PENY-) PE CORP NY.
XX PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
XX PA (SINO-) SINOGENOMAX CO LTD.
```

---

```
PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
XX WPI; 2001-355877/37.
XX N-PSDB; AAH62788.
XX
XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus
XX (WSBV), useful for producing viral polypeptides that can be used to
XX screen for agents that are useful for treating WSBV infection -
XX
XX Claim 1; Figure 3; 626pp; English.
XX
XX The invention provides the primary nucleotide sequence of the WSBV genome
XX (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
XX encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
XX (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
XX molecules and proteins of the invention are useful for diagnosis and
XX monitoring viral infection, in screens for antiviral agents and for
XX monitoring viral gene expression or activity during a treatment regimen.
XX The nucleic acid molecules are also useful as antisense constructs to
XX control viral gene expression in infected cells and tissues and to create
XX transgenic viral resistant shrimp.
XX
XX Sequence 1141 AA;
XX
XX Query Match 11.1%; Score 88.5; DB 22; Length 1141;
XX Best Local Similarity 23.3%; Pred. No. 11;
XX Matches 30; Conservative 31; Mismatches 43; Indels 25; Gaps 4;
XX
XX Qy 25 STIVSEEDFILPVYKGELEKGYQFDGWEISFGFEC--KKDAGYVNLKDTFTKPVFKKIE 82
XX Db 621 SNVEEER-----EEEQMEEEEEVEEVEEGSKEDDGDGDA-----PAQEEME 662
XX
XX Qy 83 EKKEEENKPTFDVSKKDNPOVNHSQLNEHSHRKEDLQREHSHKSDSTKDVATVLDKNN 142
XX Db 663 EKKEEEOQ-----QQPEEENGENQEEEQQQQQQPEREENKADSDSDSDSSSSSSSS 717
XX
XX Qy 143 ISSKSTNN 151
XX Db 718 SSSSSSSSS 726
XX
XX RESULT 31
XX ABP73992
XX ID ABP73992 standard; Protein; 225 AA.
XX
XX AC ABP73992;
XX
XX DT 30-JAN-2003 (first entry)
XX
XX DE Candida albicans essential protein SEQ ID NO 7829.
XX
XX KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
XX KW signal transduction; DNA replication; cell division; growth;
XX KW proliferation; Candida albicans; fungicide; antifungal.
XX
XX OS Candida albicans.
XX
XX PN WO200253728-A2.
XX
XX PD 11-JUL-2002.
XX
XX PF 26-DEC-2001; 2001WO-US49486.
XX
XX PR 29-DEC-2000; 2000US-259128P.
XX
XX PR 20-FEB-2001; 2001US-0792024.
XX
XX PR 22-AUG-2001; 2001US-314050P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX WPI; 2002-566694/60.
XX N-PSDB; ABZ32542.
```

XX Constructing strains for identifying gene products as effective targets  
PT for therapeutic intervention, by inactivating in the strain one allele  
PT of a gene and placing other allele of the gene under conditional  
PT expression -  
XX  
PS Claim 44; SEQ ID NO 7829; 167pp + Sequence Listing; English.  
XX  
XX The invention relates to constructing (M1) a strain of diploid fungal  
XX cells in which both alleles of a gene are modified, comprising modifying  
XX one allele by insertion or replacement by a cassette having an  
XX expressible selectable marker and modifying other allele by  
XX recombination, of a promoter replacement fragment with a heterologous  
XX promoter, so that expression of the second allele is regulated by the  
XX promoter. (M1) is useful for constructing a strain of diploid fungal  
XX cells in which both alleles of a gene are modified. The diploid fungal  
XX cells having both alleles modified are useful for identifying a gene that  
XX is essential to the survival or growth of a fungus, a gene that  
XX contributes to the virulence and/or pathogenicity of a fungus, a gene  
XX that contributes to the resistance of a diploid fungus to an antifungal  
XX agent, an antifungal agent that inhibits the growth of a diploid fungus  
XX and for identifying a therapeutic agent for treatment of a mammalian  
XX disease. (M1) is useful for identifying a compound which modulates the  
XX activity of a gene product, preferably enzymatic activity, carbon  
XX compound catabolism, biosynthetic, transporter, transcriptional,  
XX translational, signal transduction, DNA replication and cell division  
XX activity. The method is useful for identifying a compound having the  
XX ability to inhibit growth or proliferation of C. albicans cells and for  
XX treating infection by C. albicans. The present sequence is that of an  
XX essential Candida albicans protein used in the method of the invention.  
XX Note: The sequence data for this patent is not represented in the printed  
XX specification but is based on sequence information supplied to Derwent by  
XX the European Patent Office.  
XX  
SQ Sequence 225 AA;  
Query Match 11.0%; Score 88; DB 23; Length 225;  
Best Local Similarity 28.0%; Pred. No. 1.2;  
Matches 30; Conservative 22; Mismatches 43; Indels 12; Gaps 4;  
QY 46 YQDGEWISGFEKGKQDAGVIVNLSKDTFIKVPFKK---IEBKKEENKPTFDVSKKK-- 99  
DB 80 YDDDDDFGFGFESSNCAKXELNLSQAIKEWKQRDLIEREKLSKKKEIEIKAKS 139  
QY 100 --DNPVNHSQLNESHRKEDLOREEH--SOKSDSTKDVATVLDKNN 142  
DB 140 TIDDFVYNSKRDNHQKILSQEKFIKRDFFLK--RGTLDWRVN 184  
RESULT 32  
AAB94584  
ID AAB94584 standard; Protein; 258 AA.  
XX  
XX AAB94584;  
XX  
XX 26-JUN-2001 (first entry)  
XX  
XX Human protein sequence SEQ ID NO:15383.  
DE  
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
XX Homo sapiens.  
XX  
XX EP1074617-A2.  
XX  
XX 07-FEB-2001.  
XX  
XX 28-JUL-2000; 2000EP-0116126.  
XX  
XX 29-JUL-1999; 99JP-0248036.  
XX  
XX 27-AUG-1999; 99JP-0300253.  
XX  
XX 11-JAN-2000; 2000JP-0118776.  
XX  
XX 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.  
XX  
XX (HELI-) HELIX RES INST.  
XX  
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX  
XX Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
XX  
XX Claim 8; SEQ ID 15383; 2537pp + CD ROM; English.  
XX  
XX The present invention describes primer sets for synthesising 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesising polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 258 AA;  
Query Match 11.0%; Score 88; DB 22; Length 258;  
Best Local Similarity 26.3%; Pred. No. 1.5;  
Matches 31; Conservative 25; Mismatches 50; Indels 12; Gaps 4;  
QY 38 YKGELEKGYQFDGWEISGFEKGKQDAGY--VINLSKDTFIKVPFKKIEBKKEENKPTFDV 95  
DB 39 WEGEDEDEVDKNDWDDDDDEKKEAEVPEVKISEK---KKIAEKIKERQCKRQBEI 95  
QY 96 SKKKNPQ-----VNHSQNLNESHKEDIQREHSQKSDSTKDVATV--LDKNNITSSK 146  
DB 96 KKRLEPEEPKVLTPPEQLADKLRLKQESSDLELAKETFGVNNVTYIGIDANWPSR 153  
RESULT 33  
AAM79318  
ID AAM79318 standard; Protein; 817 AA.  
XX  
XX AAM79318;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human protein SEQ ID NO 2964.  
DE  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.  
XX  
XX Homo sapiens.  
XX  
XX WO200157190-A2.  
XX

PD 09-AUG-2001.  
XX  
PF  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
XX  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAK52451.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 215; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM8020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
XX Sequence 817 AA;  
Query Match 11.0%; Score 88; DB 22; Length 817;  
Best Local Similarity 27.3%; Pred. No. 7.4; Mismatches 55; Indels 12; Gaps 4;  
Matches 33; Conservative 21;  
QY 40 GELEKGYQFDGWEISG--FEKGDAGYVINLSKDTFKPVPFKIEEKEENKPTFDVSK 97  
Db 196 GOEKKQESFKSWEASGKHQEVSKPAVLEQRKQDTSKLRSTLPEQKKQEIISKPSQ 255  
QY 98 -KKNPQVNHSQLNESHKED-----LQREHSQKSDSTKVATVTLDKNNISSKT 148  
Db 256 WKQDTPKSKAGYVQEHKKQETPKLWPVQLQKEQ-DPKKQTPKSWTPSMQSEQNTTKSWT 314  
QY 149 T 149  
Db 315 T 315  
RESULT 34  
AAM79319  
ID AAM79319 standard; Protein; 817 AA.  
XX  
XX AAM79319;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human protein SEQ ID NO 2965.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorder; arthritis; inflammation.  
XX Homo sapiens.  
XX WO200157190-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 05-FEB-2001; 2001WO-US04098.  
XX  
XX 03-FEB-2000; 2000US-0496914.  
PR 27-APR-2000; 2000US-0560875.  
PR 20-JUN-2000; 2000US-0598075.  
PR 19-JUL-2000; 2000US-0620325.  
PR 01-SEP-2000; 2000US-0654936.  
PR 15-SEP-2000; 2000US-0663561.  
PR 20-OCT-2000; 2000US-0693325.  
PR 30-NOV-2000; 2000US-0728422.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAK52452.  
XX  
XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -  
XX  
XX Claim 20; Page 215; 6221pp; English.  
XX  
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM8020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.  
XX  
XX Sequence 817 AA;  
Query Match 11.0%; Score 88; DB 22; Length 817;  
Best Local Similarity 27.3%; Pred. No. 7.4; Mismatches 55; Indels 12; Gaps 4;  
Matches 33; Conservative 21;  
QY 40 GELEKGYQFDGWEISG--FEKGDAGYVINLSKDTFKPVPFKIEEKEENKPTFDVSK 97  
Db 196 GOEKKQESFKSWEASGKHQEVSKPAVLEQRKQDTSKLRSTLPEQKKQEIISKPSQ 255  
QY 98 -KKNPQVNHSQLNESHKED-----LQREHSQKSDSTKVATVTLDKNNISSKT 148  
Db 256 WKQDTPKSKAGYVQEHKKQETPKLWPVQLQKEQ-DPKKQTPKSWTPSMQSEQNTTKSWT 314  
QY 149 T 149  
Db 315 T 315  
RESULT 35  
ABP73209  
ID ABP73209 standard; Protein; 635 AA.  
XX  
XX ABP73209;  
AC



the invention are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Such conditions include cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell disorders), chronic inflammatory conditions (e.g., asthma or arthritis), proliferative retinopathy, atherosclerosis, coronary heart disease, arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

Sequence 2515 AA;

Query Match 11.0%; Score 87.5; DB 22; Length 2515;

Best Local Similarity 24.2%; Pred. No. 40;

Matches 38; Conservative 29; Mismatches 55; Indels 35; Gaps 6;

QY 5 VSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGY 64

DB 2352 IAELOREPLLVNESLNVENSGFRTEEIHSESYNGKISGRKDAEISGHSVEADP-- 2409

QY 65 VINLSKDTFIKPVFKKIEEKEENKPTFDVSKK-----DNPQVNHSQLNESH 113

DB 2410 -----KEVEE--EERHMPK---RKRQHYLSSEDEPDNDPDLDSRI-ETA 2449

QY 114 RKEDLOREHQSQSDTKDVTATVLDKNNISKSTTN 150

DB 2450 QRCQPEPHATKEENSRLDEE--LPKTSSETNSTTS 2484

#### RESULT 37

AAM80268

ID AAM80268 standard; Protein; 2515 AA.

XX AC AAM80268;

DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 3914.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663561.

XX PR 20-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0728422.

XX PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
XX WPI; 2001-476283/51.  
DR N-PSDB; AAKS3401.

XX Nucleic acids encoding polypeptides with cytokine-like activities,  
PT useful in diagnosis and gene therapy -

PS Claim 20; Page 465-466; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the  
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
CC (AAM80020) are omitted as the relevant pages from the sequence listing  
CC were missing at the time of publication.

SQ Sequence 2515 AA;

Query Match 11.0%; Score 87.5; DB 22; Length 2515;

Best Local Similarity 24.2%; Pred. No. 40;

Matches 38; Conservative 29; Mismatches 55; Indels 35; Gaps 6;

QY 5 VSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGY 64

DB 2352 IAELOREPLLVNESLNVENSGFRTEEIHSESYNGKISGRKDAEISGHSVEADP-- 2409

QY 65 VINLSKDTFIKPVFKKIEEKEENKPTFDVSKK-----DNPQVNHSQLNESH 113

DB 2410 -----KEVEE--EERHMPK---RKRQHYLSSEDEPDNDPDLDSRI-ETA 2449

QY 114 RKEDLOREHQSQSDTKDVTATVLDKNNISKSTTN 150

DB 2450 QRCQPEPHATKEENSRLDEE--LPKTSSETNSTTS 2484

#### RESULT 38

AAM80269

ID AAM80269 standard; Protein; 2515 AA.

XX AC AAM80269;

DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 3915.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.



```
RESULT 40
AAG47057
ID AAG47057 standard; Protein; 313 AA.
XX
AC AAG47057;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 59270.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147152.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
```



PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 10.9%; Score 87; DB 21; Length 313;  
Best Local Similarity 25.4%; Pred. No. 2.5;  
Matches 36; Conservative 26; Mismatches 46; Indels 34; Gaps 6;  
QY 6 SELKPHRVTVTQNGKMSSTIVSEEDFILPVYKGELEKGYFDGWEISG-FEGKRDAGY 64  
Db 55 SSEKPNR--KXIQKGKIKSS-----PADG-KLSGKMKRKEKVG 91  
QY 65 VINLSKDTIKPVFKIEKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLOREHS 124  
Db 92 NYVDISEPILEAISTEKVKKGGMMNK-----TKKRAEIEITRSSV-----EDLKRESKF 141  
QY 125 OKSDSTKDVATVLDKNNISSK 146  
Db 142 KKSNNKKKMDMTSKKENKIEEE 163

Search completed: February 10, 2004, 10:53:55  
Job time : 37.1775 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 29.0366 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_610\_773

Perfect score: 848

Sequence: 1 TTVKEFLNKTGVESELKP.....ATVLDKNISSKSTNNPNK 164

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archheap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	848	100.0	2119	2 Q9AHT5	Q9AHT5 streptococc
2	848	100.0	2140	16 Q97RY6	Q97RY6 streptococc
3	845	99.6	2144	16 Q8DQP7	Q8DQP7 streptococc
4	844	99.5	2144	2 Q9S4M8	Q9S4M8 streptococc
5	119	14.0	775	16 Q8CPK8	Q8CPK8 staphylococ
6	112.5	13.3	361	5 Q9SP15	Q9SP15 plasmodium
7	112.5	13.3	379	5 Q9U6C4	Q9U6C4 plasmodium
8	111.5	13.1	346	5 Q9U0G0	Q9U0G0 plasmodium
9	111.5	13.1	354	5 Q25995	Q25995 plasmodium
10	111.5	13.1	354	5 Q81J55	Q81J55 plasmodium
11	111.5	13.1	379	5 Q25706	Q25706 plasmodium
12	111	13.1	1038	13 Q90784	Q90784 gallus gall
13	110.5	13.0	379	5 Q25705	Q25705 plasmodium
14	110	13.0	829	5 Q815F3	Q815F3 plasmodium
15	109.5	12.9	609	5 Q812K8	Q812K8 plasmodium
16	109	12.9	380	5 Q26019	Q26019 plasmodium

17	109	12.9	3008	5	Q81436	Q81436 plasmodium
18	108.5	12.8	600	5	Q77355	Q77355 plasmodium
19	107.5	12.7	470	10	Q9FJK9	Q9FJK9 arabidopsis
20	106	12.5	951	5	Q96229	Q96229 plasmodium
21	104.5	12.3	329	5	Q9NFV9	Q9NFV9 plasmodium
22	104	12.3	540	10	Q94C59	Q94C59 arabidopsis
23	104	12.3	540	10	Q9SAB4	Q9SAB4 arabidopsis
24	103.5	12.2	325	5	O44016	O44016 dictyosteli
25	103	12.1	2081	10	Q9LH98	Q9LH98 arabidopsis
26	102.5	12.1	238	5	Q81226	Q81226 plasmodium
27	101.5	12.0	312	16	Q9PPL5	Q9PPL5 campylobact
28	101.5	12.0	382	5	Q9V7J0	Q9V7J0 drosophila
29	101.5	12.0	556	5	Q9V7J9	Q9V7J9 drosophila
30	101.5	12.0	785	5	Q9GQ82	Q9GQ82 drosophila
31	101.5	12.0	948	5	Q9U4U6	Q9U4U6 plasmodium
32	101.5	12.0	1377	5	Q81586	Q81586 plasmodium
33	100.5	11.9	1373	5	Q81353	Q81353 plasmodium
34	100	11.8	375	4	O14712	O14712 homo sapien
35	100	11.8	622	16	Q98QA1	Q98QA1 mycoplasma
36	100	11.8	1130	5	Q81JZ4	Q81JZ4 plasmodium
37	100	11.8	3127	5	Q81DA0	Q81DA0 plasmodium
38	99.5	11.7	385	5	Q93424	Q93424 caenorhabdi
39	99.5	11.7	4524	5	Q813J9	Q813J9 plasmodium
40	99	11.7	211	5	P91488	P91488 caenorhabdi
41	99	11.7	219	16	Q9XOM6	Q9XOM6 thermotoga
42	99	11.7	1859	5	Q81C27	Q81C27 plasmodium
43	99	11.7	2162	5	Q81BH2	Q81BH2 plasmodium
44	99	11.7	2563	5	Q813A0	Q813A0 plasmodium
45	99	11.7	3026	5	Q81LS9	Q81LS9 plasmodium

#### ALIGNMENTS

RESULT 1

Q9AHT5 PRELIMINARY; PRT; 2119 AA.

AC Q9AHT5; 01-JUN-2001 (Tremblrel. 17, Created)

DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DE Serine protease (Fragment).

GN PRTA.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

CC Streptococcus.

OX NCBI\_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=N4;

RX MEDLINE=21116976; PubMed=11179332;

RA Wisemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,

RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,

RA Gayle A., Brewash Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,

RA Langermann S., Johnson S., Koenig S.;

RT "Use of a Whole Genome Approach To Identify Vaccine Molecules

RT Affording Protection against Streptococcus pneumoniae Infection.";

RL Infect. Immun. 69:1593-1598 (2001).

CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY

EMBL; AF291699; AAK19159.1; -.

DR HSSP; P00782; 2SPT.

DR MEROPS; S08.064; -.

DR InterPro; IPR001899; Gram\_pos\_anchor.

DR InterPro; IPR006192; LPXTG.

DR InterPro; IPR003137; PA.

DR InterPro; IPR000209; Peptidase\_S8.

DR InterPro; IPR001680; WD40.

DR Pfam; PF00746; Gram\_pos\_anchor; 1.

DR Pfam; PF02225; PA; 1.

DR Pfam; PF00082; Peptidase\_S8; 2.

DR PRINTS; PR00723; SUBTILISIN.

DR TIGRFAMs; TIGR01167; LPXTG\_anchor; 1.

```

DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00840; PA; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 1.
KW Cell wall; Peptidoglycan-anchor; Protease.
FT NON_TER ... 1
SQ SEQUENCE 2119 AA; 238226 MW; 517F9B7F6B960A6A CRC64;

  Query Match      100.0%; Score 848; DB 2; Length 2119;
  Best Local Similarity 100.0%; Pred. No. 9,6e-53;
  Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKFEILNKDTGEVSELKPHRVVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60
Db 1922 TTVKFEILNKDTGEVSELKPHRVVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDG 1981

QY 61 WEISGFEKKDAGYVILNLSKDTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLN 120
Db 1982 WEISGFEKKDAGYVILNLSKDTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLN 2041

QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164
Db 2042 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2085

  RESULT 2
  Q97RY6 PRELIMINARY; PRT; 2140 AA.
  AC Q97RY6
  DT 01-OCT-2001 (TREMBlrel. 18, Created)
  DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
  DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
  DE Serine protease, subtilase family.
  GN SP0641
  OS Streptococcus pneumoniae.
  OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
  OC Streptococcus.
  OX NCBI_TaxID=1313;
  RN [1]
  RP SEQUENCE FROM N.A.
  RX MEDLINE=21357209; PubMed=1146316;
  RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
  RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
  RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
  RA Umavam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
  RA Holzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
  RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
  RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
  RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
  RT "Complete genome sequence of a virulent isolate of Streptococcus
  RT pneumoniae";
  RL Science 293:498-506 (2001).
  DR EMBL; AE007373; AAK74791.1; -.
  DR MEROPS; S08.064; -.
  DR TIGR; SP0641; -.
  DR InterPro; IPR001899; Gram_pos_anchor.
  DR InterPro; IPR006192; LPXTG.
  DR InterPro; IPR003137; PA.
  DR InterPro; IPR000209; Peptidase_S8.
  DR InterPro; IPR001680; WD40.
  DR Pfam; PF00746; Gram_pos_anchor; 1.
  DR Pfam; PF02225; PA; 1.
  DR Pfam; PF00082; Peptidase_S8; 2.
  DR PRINTS; PR00723; SUBTILISIN.
  DR TIGRfam; TIGR01167; LPXTG_anchor; 1.
  DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
  DR PROSITE; PS00840; PA; 1.
  DR PROSITE; PS00137; SUBTILASE_HIS; 1.
  DR PROSITE; PS00138; SUBTILASE_SER; 1.
  DR PROSITE; PS00678; WD_REPEATS_1; 1.
  KW Protease; Complete proteome.

SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;

  Query Match      100.0%; Score 848; DB 16; Length 2140;
  Best Local Similarity 100.0%; Pred. No. 9,7e-53;
  Matches 164; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKFEILNKDTGEVSELKPHRVVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60
Db 1943 TTVKFEILNKDTGEVSELKPHRVVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDG 2002

QY 61 WEISGFEKKDAGYVILNLSKDTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLN 120
Db 2003 WEISGFEKKDAGYVILNLSKDTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLN 2062

QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164
Db 2063 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2106

  RESULT 3
  Q8DQP7 PRELIMINARY; PRT; 2144 AA.
  AC Q8DQP7
  DT 01-MAR-2003 (TREMBlrel. 23, Created)
  DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
  DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
  DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
  GN PrtA OR SP0561.
  OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
  OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
  OC Streptococcus.
  OX NCBI_TaxID=171101;
  RN [1]
  RP SEQUENCE FROM N.A.
  RX MEDLINE=21429245; PubMed=11544234;
  RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., BURGESS S.,
  RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer S.,
  RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
  RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
  RA McAhren S.M., McHenry M., Mcleaster K., Mundy C.W., Nicas T.I.,
  RA Norris P.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
  RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
  RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
  RA Glass J.I.;
  RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
  RL J. Bacteriol. 183:5709-5717 (2001).
  DR EMBL; AE008434; AAK9365.1; -.
  KW Hydrolase; Complete proteome.
  SQ SEQUENCE 2144 AA; 240436 MW; 8C1B4B1DBC503A0C CRC64;

  Query Match      99.6%; Score 845; DB 16; Length 2144;
  Best Local Similarity 99.4%; Pred. No. 1,6e-52;
  Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTVKFEILNKDTGEVSELKPHRVVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60
Db 1947 TTVKFEILNKDTGEVSELKPHRVVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDG 2006

QY 61 WEISGFEKKDAGYVILNLSKDTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLN 120
Db 2007 WEISGFEKKDAGYVILNLSKDTFIKPVFKKIEKKEENKPTFDVSKKDNQVNHSQLN 2066

QY 121 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 164
Db 2067 ESHRKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 2110

  RESULT 4
  Q9S4M8 PRELIMINARY; PRT; 2144 AA.
  AC Q9S4M8
  DT 01-MAY-2000 (TREMBlrel. 13, Created)
  DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
  KW Protease; Complete proteome.

```

01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
Cell wall-associated serine proteinase PrtA precursor.  
PRTA.  
Streptococcus pneumoniae.  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN (1)  
RC STRAIN=3.B;  
SEQUENCE FROM N.A.  
RA Bethe G., ten Thoren E., Bongaerts R.J.M., Heinz H.-P., Zysk G.;  
RT "Cloning and sequencing of a novel surface protease of Streptococcus  
pneumoniae."  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
AN AMIDE BOND (BY SIMILARITY).  
DR EMBL; AF127143; AAD48399.1; -;  
DR HSSP; P00782; 2SPT.  
DR MEROPS; S08.064; -;  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF0746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRfams; TIGR01167; LPXTG\_anchor; 1.  
DR PROSITE; PS50847; GRAM\_POS\_ANCHORING; 1.  
DR PROSITE; PS50840; PA; 1.  
DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
KW Cell wall; Peptidoglycan-anchor; Signal.  
FT SIGNAL 1 19  
FT CHAIN 20 2144  
FT CDS 20 2144  
FT CELL WALL-ASSOCIATED SERINE PROTEINASE  
FT PRTA.  
SQ SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;  
  
Query Match 99.5%; Score 844; DB 2; Length 2144;  
Best Local Similarity 98.8%; Pred. No. 1.9e-52;  
Matches 162; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 TTVEFLIKDGTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYFDG 60  
Db 1947 TTVEFLIKDGTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYFDG 2006  
  
Qy 61 WEISGFEKGDAGYVNLKDTFKIPVFKKIEEKEENKPTFDVSKKONPQVNHSQLN 120  
Db 2007 WEISGFEKGDAGYVNLKDTFKIPVFKKIEEKEENKPTFDVSKKONPQVNHSQLN 2066  
  
Qy 121 ESHRKEDLQREHKSQSDTKDVTATVLDKNNISSKSTNNPNK 164  
Db 2067 ESHRKEDLQREHKSQSDTKDVTATVLDKNNISSKSTNNPNK 2110  
  
RESULT 5  
Q8CPK8 ID Q8CPK8 PRELIMINARY; PRT; 775 AA.  
AC Q8CPK8;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Penicillin-binding protein 1.  
GN SE0856.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1282;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12228;  
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,

RA Chen Z., Wen Y.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE016746; AAO04453.1; -;  
KW Complete proteome.  
SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7B98777C CRC64;  
  
Query Match 14.0%; Score 119; DB 16; Length 775;  
Best Local Similarity 27.0%; Pred. No. 1.1;  
Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
  
Qy 10 KDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYK-----CELEKGYQFDG-- 61  
Db 614 EDSVNAQSLKP-----ITGNGKQIKQOSVKSGTKVLPHSKVLMGTDELTPW-DMTGWTK 668  
  
Qy 62 -EISGPF-----GKKDAGYVNL--LSKDTFKIPVFKKIEEKEENKPTFDVS----K 107  
Db 669 EDVLAFEDLTIKIVSTYKNGFGVTVNQSIKQILK-----NKDKIEVLSNAED 715  
  
Qy 108 KCONPQVNHSQLNESHKEDLQREHKSQSDTKDVTATVLDKNNISSKSTNN 161  
Db 716 TDDQEKTDSDSDNKSCKDADEHSHNTSSSTKN-----DKSNADSKNDSDD 763  
  
RESULT 6  
Q95PI5 ID Q95PI5 PRELIMINARY; PRT; 361 AA.  
AC Q95PI5;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Merozoite surface protein 3 (fragment).  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVO;  
RA Hisaeda H., Saul A., Long C.A., Miller L.H., Stowers A.W.;  
RT "Merozoite Surface Protein 3 and Protection Against Malaria in Aotus  
Monkeys."  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY044180; AAK94780.1; -;  
FT NON TER 1 1  
FT NON TER 361 361  
SQ SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;  
  
Query Match 13.3%; Score 112.5; DB 5; Length 361;  
Best Local Similarity 22.4%; Pred. No. 1.4;  
Matches 43; Conservative 33; Mismatches 55; Indels 61; Gaps 8;  
  
Qy 11 DTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQ----- 57  
Db 147 ETGE-----RNSRNNFYTKTKE-----YAKVEKDYERAKYAKQAV 187  
  
Qy 58 -----FD---GWEISGF---EGKKDAG-----YVNLKDTFKIPVFKKIEEKEEN 99  
Db 188 LKAKEASSYDILGWEPGGVPEHKKEENMLSHLYVSSKDKENISKENDVDLDE-KEEA 246  
  
Qy 100 KPTFDVSKKONPQVNHSQLNESHKEDLQREHKSQSDTKDVTATVLDKNN----- 151  
Db 247 EETEEELKEEKEEETEEISEDEEEEEEEEEEEEEEEEEEEEEEEEEEEENNDOKDME 306  
  
Qy 152 --NISSKSTNN 161  
Db 307 AQNLISKQNNN 318  
  
RESULT 7  
Q9U6C4 ID Q9U6C4 PRELIMINARY; PRT; 379 AA.  
AC Q9U6C4;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)



RA Perteza M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrell B.,  
RT "Genome sequence of the human malaria parasite Plasmodium  
RT falciparum";  
RL Nature 419:498-511(2002).  
DR EMBL; AE014834; AAN35542.1; -.  
SQ SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

Query Match 13.1%; Score 111.5; DB 5; Length 354;  
Best Local Similarity 22.5%; Pred. No. 1.7;  
Matches 41; Conservative 34; Mismatches 62; Indels 45; Gaps 7;

QY 1 TTIVKEFILN-KDTGVSSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFD 59  
DB 154 TKTEYAEAKNAYEKANQANQANQAVLKAEASS-----YDYL----- 194

QY 60 GWEISGF--EGKQDAG-----YVINLSKDTFTKPVFKKIEKKEE-----EN 99  
DB 195 GWEFGGVPHEKKENMLSHLYVSSKDKENISKENDVDLDEKEEAEETEEELKNEE 254

QY 100 KPTFDVSKKKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTVATVLDKNNISSKSTT 159  
DB 255 ETESEISDEEBEEREEBENDKKQEKEQSNENNDDQKDMA-----QNLISKQNN 309

QY 160 NN 161  
DB 310 NN 311

RESULT 11  
Q25706  
ID Q25706 PRELIMINARY; PRT; 379 AA.  
AC Q25706;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Polymorphic antigen.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C812;  
RX MEDLINE=98156743; PubMed=9497029;  
RA McColl D.J., Anders R.F.;  
RT "Conservation of structural motifs and antigenic diversity in the  
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
RL Mol. Biochem. Parasitol. 90:21-31(1997).  
DR EMBL; U08852; AAC47832.1; -.  
SQ SEQUENCE 379 AA; 43302 MW; ABF9D54E1ED91A24 CRC64;

Query Match 13.1%; Score 111.5; DB 5; Length 379;  
Best Local Similarity 23.6%; Pred. No. 1.8;  
Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8;

QY 19 KPHRVTVTIQNGKMSSTIVSEEDF-----ILPVYKGELEKGYQFD-GWEISGF- 66  
DB 171 KPSRLNLSRKTKEAEQV--EKDYERAKNAYQKANAQAVLKAEASSYDYLWGFGGV 228

QY 67 -EGKQDAG-----YVINLSKDTFTKPVFKKIEKKEEKPTFDVSKKKNPQVNHSQLN 120  
DB 229 PEHKKENMLSHLYVSSKDKENISKENDVDLDE-KEEAEETEEELKNEEETESIS 287

QY 121 -----ESHKEDLQREHSQKSDSTKDVTVATVLDKNNISSKSTNN 161  
DB 288 EDEEBEEREEBENDKKQEKEQSNENNDDQKDMA-----QNLISKQNNN 336

Query Match 13.0%; Score 110.5; DB 5; Length 379;  
Best Local Similarity 22.1%; Pred. No. 2.1;  
Matches 43; Conservative 35; Mismatches 48; Indels 69; Gaps 9;

ID Q90784 PRELIMINARY; PRT; 1038 AA.  
AC Q90784;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Claustrian.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Cole G.J.;  
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-451 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=94157526; PubMed=7906711;  
RA Burg M.A., Cole G.J.;  
RT "Claustrian, an antiadhesive neural keratan sulfate proteoglycan, is  
RT structurally related to MAP1B.";  
RL J. Neurobiol. 25:1-22(1994).  
DR EMBL; X67778; CAA47988.1; -.  
SQ SEQUENCE 1038 AA; 117112 MW; 213D694A5B510927 CRC64;

Query Match 13.1%; Score 111; DB 13; Length 1038;  
Best Local Similarity 23.7%; Pred. No. 5.9;  
Matches 44; Conservative 28; Mismatches 58; Indels 56; Gaps 5;

QY 3 VKFELNKDTGVSSELKPHRVTVTIQ----- 29  
DB 528 VKQAKLKQRTDSKESLAPAKTTTKQDCQKRNKKKHSLSLVQOLEKPKQLESKEKTP 587

QY 30 -----GKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEGKKDAGYVINLSKDTF 83  
DB 588 VKKEKAVKPKETIVAEDKDV-----TTKEQLGKSETSEKQASEKQDVKPKVTKES 639

QY 84 IKPVFK-KIEKKEENKPTFDVSKKKNPQVNHSQLNESHKEDLQRE-----EHSQKS 137  
DB 640 VKKEKAVKPKETIVAEDKDV-----TTKEQLGKSETSEKQASEKQDVKPKVTKES 696

QY 138 DSTKDV 143  
DB 597 EAKKEV 702

RESULT 13  
Q25705  
ID Q25705 PRELIMINARY; PRT; 379 AA.  
AC Q25705;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Polymorphic antigen.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K1;  
RX MEDLINE=98156743; PubMed=9497029;  
RA McColl D.J., Anders R.F.;  
RT "Conservation of structural motifs and antigenic diversity in the  
RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
RL Mol. Biochem. Parasitol. 90:21-31(1997).  
DR EMBL; U08851; AAC47831.1; -.  
SQ SEQUENCE 379 AA; 43344 MW; DC7AF106887C8AA0 CRC64;

Query Match 13.0%; Score 110.5; DB 5; Length 379;  
Best Local Similarity 22.1%; Pred. No. 2.1;  
Matches 43; Conservative 35; Mismatches 48; Indels 69; Gaps 9;

```

QY 11 DTGEVSELKPHRVTVTQNGKEMSTIVSEEDFILPVYKGELEKGYQ-----57
Db 167 ETGCE-----RNSRNNFYTTKTE-----YAKVEKDYERAKNAYQKANOAV 207
QY 58 -----FD--GWEISGF--EGKKDAG-----YVINLSKDTFTKVPFKKIEEKKEEN 99
Db 208 LKAKEASSYDILGWFEGGVPEHKKEENLHLYVSSKDKENISKENDDVIDE-KEEA 266
QY 100 KPTFDVSKKNDPQNVHNSLN-----ESHKEDLQREHSQKSDSTKDVAT 146
Db 267 BETESEELEKNEETESISEDEEEEEEKKEKQEQSQNNNDOKDMA- 325
QY 147 VLDKNNISSKSTNN 161
Db 326 ----QNLISKNNNN 336

RESULT 14
Q815F3 ID Q815F3 PRELIMINARY; PRT; 829 AA.
AC Q815F3;
DT 01-WAR-2003 (TrEMBLrel. 23, Created)
DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Hypothetical protein.
GN PFL1275C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=2255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum."
RL Nature 419:498-511(2002).
DR EMBL; A5014848; AAN36341.1; --
KW Hypothetical protein.
SQ SEQUENCE 829 AA; 98816 MW; EF2675E301B2CE93 CRC64;

Query Match 13.0%; Score 110; DB 5; Length 829;
Best Local Similarity 24.0%; Pred. No. 5.4;
Matches 43; Conservative 39; Mismatches 55; Indels 42; Gaps 9;

QY 2 TVKEFLNKDGEVSELKPHRVTVTQNGKEMSTIVSEEDFILPVYKGELEKGYQFDGW 61
Db 491 TSDNSILNTNKKI--IQP--LEYLKNIGD--KTLMTERDVLDFVHPYMIKKYHLN-- 542
QY 62 EISGFGKDGAGVINLSKDTFTKVPFKKIEEKKEENKPT-----FDVSK 107
Db 543 -----KKETLNFNSLN-----FREIEKNRDKKKGTHNNKNDAEYWLKYIKK 587
QY 108 KKDN-PQNVHNSLNESH-KREDLQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164
Db 588 KKKNFENNNTLSDNSNKKNNKLVEH---DNSLKQEQIINDKNVIEHTKIYDNOKK 643

RESULT 15
Q812K8 ID Q812K8 PRELIMINARY; PRT; 609 AA.
AC Q812K8;
DT 01-WAR-2003 (TrEMBLrel. 23, Created)
DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
```

```

DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Large cyclophilin-like protein.
GN PFI1490C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Gobie A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrall B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL929358; CAD51984.1; --
SQ SEQUENCE 609 AA; 72552 MW; 8CDF86E85FF9A021 CRC64;

Query Match 12.9%; Score 109.5; DB 5; Length 609;
Best Local Similarity 23.9%; Pred. No. 4.2;
Matches 42; Conservative 36; Mismatches 79; Indels 19; Gaps 5;

QY 5 EPIFNKDTGEVSELKPHRVTVTQNGKEMSTI-----VSEEDFIL---PVYKGELEKGYQ 57
Db 114 KIILNRKTNVYSSEIYKL---FHNKNEFDPITHDKISKEDFVLQDPLNNKTSYVASH 170
QY 58 FDGWEISGFGKDGAGVINLSKDTFTKVPFKKIEEKKEENKPTFDVSKKNDP----- 112
Db 171 LNDRDQEDNEKKKNIFAPSIQDNGAIMSILKMEAEKKEKKEEKKKKYVIL 230
QY 113 QVNSHLSNESHKEDLQREHSQKSDSTKDVATVLD-----KNNISSKSTNNPNK 164
Db 231 QDYNDSNESHSEHSNKKKMLKNKKDEYKNKIHSTNYDSDSSENEIKICTNYSDNK 286

RESULT 16
Q26019 ID Q26019 PRELIMINARY; PRT; 380 AA.
AC Q26019;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Polymorphic antigen precursor.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RX MEDLINE=95198774; PubMed=7891748;
RA McColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
RA Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT "Molecular variation in a novel polymorphic antigen associated with
Plasmodium falciparum merozoites."
RL Mol. Biochem. Parasitol. 68:53-67(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FC27;
RX MEDLINE=98156743; PubMed=9497029;
RA McColl D.J., Anders R.F.;
RT "Conservation of structural motifs and antigenic diversity in the
Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
```



RL Mol. Biochem. Parasitol. 90:21-31(1997).  
 DR EMBL; L07944; AAC09378.1; -  
 KW SIGNAL.  
 FT CHAIN.  
 FT SIGNAL.  
 FT CHAIN.  
 SQ SEQUENCE 380 AA; 43290 MW; 0986CAL1393094CA2 CRC64;  
 Query Match 12.9%; Score 109; DB 5; Length 380;  
 Best Local Similarity 23.5%; Pred. No. 2.7;  
 Matches 42; Conservative 34; Mismatches 67; Indels 36; Gaps 8;  
 QY 12 TGEVSELPKRVTVTIQNGKMSSTIVSEEPFILPVY-----KGELEKGYQFD-GWE 62  
 DB 166 TSETPE-KPSRINLPSRTKYEAKAKNAYQKANOVLKAKEASSYDYLWGE 224  
 QY 63 ISGF--EGKKDAG-----YVINLSKDTFTPKVFKIEEKEENKPTFDVSKKDNPOVN 115  
 DB 225 FGGGVPEKKKENMLSHYVSKDKENISKENDVDLDE-KEEAEETEELKEENKEET 283  
 QY 116 HSQLN-----ESHKEDLOREHSQKSDTKDVTATVLDKNNISSKSTTN 161  
 DB 284 ESEISEDEEBEKEENKKEKEQKQENNNQKDMEA-----QNLISKNNQNN 337  
 RESULT 17  
 ID Q81436 PRELIMINARY; PRT; 3008 AA.  
 AC Q81436;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DE Hypothetical protein.  
 GN PFE0325W.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RP SEQUENCE FROM N.A.  
 RA Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,  
 RA Hall N., Bowman S., Churcher C., Quail M., Barrell B.;  
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN  
 RX MEDLINE=22255708; PubMed=12368867;  
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
 RA Cronin A., Davies R., Davis P., Dear P., Dearn F., Doggett J.,  
 RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphray S., Jagals K., James K.D., Johnson D., Kerhornou A.,  
 RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,  
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,  
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;  
 RA "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";  
 RL Nature 419:527-531(2002).  
 DR EMBL; AL929351; CAD51431.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 3008 AA; 356023 MW; 60CBEBE15C599B4 CRC64;

Query Match 12.9%; Score 109; DB 5; Length 3008;  
 Best Local Similarity 32.4%; Pred. No. 26;  
 Matches 35; Conservative 22; Mismatches 31; Indels 20; Gaps 6;  
 QY 74 YVINLSK-----DTFTKPVFKIEEKEENKPTFDVSKKDNPOVNHQ----LNEHSHKE 126  
 DB 2310 YDIELSKIEKFGASIGPVFTD-EENKEENK--EVNKEENKKEENKQEVNKE 2366

QY 127 DLQREH-----SQKSDTKDVTATVLDKNNISSK-----STTNPNK 164  
 DB 2367 ENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENKKEENK 2414  
 RESULT 18  
 ID O77355 PRELIMINARY; PRT; 600 AA.  
 AC O77355;  
 DT 01-NOV-1998 (TRENBLrel. 08, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE Hypothetical 71.7 Kda protein.  
 GN PFC0465C, MAL3P4.20.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=99376085; PubMed=10448855;  
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,  
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,  
 RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,  
 RA Horrocks P., Jagals K., Jassal B., Kyes S., McLean J., Moule S.,  
 RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,  
 RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,  
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;  
 RA "The complete nucleotide sequence of chromosome 3 of Plasmodium  
 RT falciparum.";  
 RL Nature 400:532-538(1999).  
 DR EMBL; AL008970; CAA15610.2; -  
 DR InterPro; IPR02483; PWI.  
 DR Pfam; PF01480; PWI; 1.  
 DR SMART; SM00311; PWI; 1.  
 DR Hypothetical protein.  
 SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;  
 Query Match 12.8%; Score 108.5; DB 5; Length 600;  
 Best Local Similarity 29.3%; Pred. No. 4.9;  
 Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;  
 QY 37 IVSEEDFILPVY-----KGELEKGYQFDGWEISGFEGKK-----DAGYVINLSKDTFTKPV 87  
 DB 60 ILGFEDDILVEYCYSQLKQSKKK--DGEDKYLNAKGLKINLTGFIGNKKSDIFIEL 116  
 QY 88 FKKI--BEKKEE-----ENKPTFDVSK-KKONPVNHSOLNE-----SHRK 125  
 DB 117 LELLNEKEKEEHTADTLNENK-TNDIKVKKNENINENYNNENKDISNKDEHVSQHN 175  
 QY 126 E-----DLQREH-----SQKSDTK-----DVTATVLDKNNISSKSTTN 160  
 DB 176 EHNINNNVLKKEEYTDIQDKRKHSLSQKSDSYKKRPFNKRTKSIER-SLSNKRYDE 234  
 QY 161 NPNK 164  
 DB 235 KTNK 238  
 RESULT 19  
 ID O9FJK9 PRELIMINARY; PRT; 470 AA.  
 AC O9FJK9;  
 DT 01-MAR-2001 (TRENBLrel. 16, Created)  
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE Gb|AAFP20218.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RX MEDLINE=99087489; PubMed=9872454;  
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
RT physically assigned P1 and TAC clones.";  
RL DNA Res. 5:297-308(1998).  
DR EMBL; AB015468; BAB10694.1; --  
SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72B35AC54 CRC64;

Query Match 12.7%; Score 107.5; DB 10; Length 470;  
Best Local Similarity 20.1%; Pred. No. 4.4;  
Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

QY 19 KHRVTVTIQTGKMSSTIVSEDFILPVYKGELEKGYQFDGWEISGFE-----GKK 70  
DB 82 RENRVTDTVNNSGSK-----YVQDLARRIRYDE-EATGSQAQRIDHPNOK 129  
QY 71 DAGYVNLKDTFTKPVFKKIEEKEENKPTFDVSKKDN-----111  
DB 130 NVGITEKAFENGPIETSHRDNDKRNQKNTAAKSSENAVSRVSFGADHKRAEVMGK 189  
QY 112 PQVHSHQLNE-----SHRKDLORREHSQKSDTKDVTATVLDKNNISKSTTNNPNK 164  
DB 190 PMENRDQVRQTESAEKSHRKNVTSEKPRDQGVKKTAKDKDRNKEKEEKTESINK 248

RESULT 20  
O96229 ID O96229 PRELIMINARY; PRT; 951 AA.  
AC O96229;  
DT 01-MAY-1999 (TremBLrel. 10, Created)  
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN FFB0680W.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=99021743; PubMed=9804551;  
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,  
RA Koonin E.V., Shalton S., Mason T., Yu K., Fujii C., Pederson J.,  
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perteau M.,  
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,  
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;  
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium  
RT falciparum.";  
RL Science 282:1126-1132(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=22255705; PubMed=1236864;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
RA Perteau M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrall B.;  
RT "Genome sequence of the human malaria parasite Plasmodium  
RT falciparum.";  
RL Nature 419:498-511(2002).  
RL EMBL; AB001410; AAC71925.2; --  
DR Hypothetical protein.  
KW  
SQ SEQUENCE 951 AA; 112486 MW; AC6D889358A84F4F CRC64;

Query Match 12.5%; Score 106; DB 5; Length 951;  
Best Local Similarity 24.3%; Pred. No. 12;  
Matches 42; Conservative 35; Mismatches 40; Indels 56; Gaps 10;

QY 10 KDTGVSBLKPHRVT-VTIQNGKMSSTIVSEDFILPVYKGELEKGYQFDGWEI--SGF 66  
DB 128 EEKNKINKSLHRLQNELNLQSGK-----NEQDI-----NKNEKGKQ----DISNSNA 170  
QY 67 EGKQDAGVYVNLKDTFTKPVFKKIEEKE-----ENKPTFD-----VSKKKNP 112  
DB 171 ENKKD-----VKEGVKELEKEEKEEKISDDHKVVEENKKSDDHKVVEENKKSDDH 218  
QY 113 QVNHSHQLNESHKEDLQR-EHHSQKSDTKDVTATVLDKNNISKSTTNNPNK 164  
DB 219 KVEENKSDDKHKBIEVKKVEEHEDEE-----DKKEKKSNNKNDENK 262

RESULT 21  
O9NFV9 ID O9NFV9 PRELIMINARY; PRT; 329 AA.  
AC O9NFV9;  
DT 01-OCT-2000 (TremBLrel. 15, Created)  
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)  
DE Merozoite surface protein 3 (Fragment).  
DE MSP3.  
GN  
OS Plasmodium falciparum (isolate 7G8).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=57266;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=7G8;  
RX MEDLINE=20416497; PubMed=10960178;  
RA Okenu D.M.N., Thomas A.W., Conway D.J.;  
RT "Allelic lineages of the merozoite surface protein 3 gene in  
RT Plasmodium reichenowi and Plasmodium falciparum.";  
RL Mol. Biochem. Parasitol. 109:185-188(2000).  
DR EMBL; AJ252287; CAB85901.1; --  
FT NON\_TER 1  
FT NON\_TER 329  
SQ SEQUENCE 329 AA; 36916 MW; C5B045DB5E21A159 CRC64;

Query Match 12.3%; Score 104.5; DB 5; Length 329;  
Best Local Similarity 24.3%; Pred. No. 4.9;  
Matches 43; Conservative 30; Mismatches 63; Indels 41; Gaps 8;

QY 1 TTVKEFILN-KDTGVSBLKPHRVTVTIQTGKMSSTIVSEDFILPVYKGELEKGYQFD 59  
DB 144 TKTKEYAEKKNAYEAKNAQKNAQAVLKAKEASS-----YDYL-----184  
QY 60 GWEISGF--EGKQDAG-----YVNLKSDTKFTKPVFKKIEEKEENKPTFDVSKKKNP 112  
DB 185 GWFGGVPVPHKKENMLSHLYVSKKENISKENDVDLDE-KEEAEETEEELKEKNE 243  
QY 113 QVNHSHQLNESHKEDLQR-EH-----SQKSDTKDVTATVLDKNNISKSTTNN 161  
DB 244 ETESEISEDEEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEEKEE 295

RESULT 22  
O94C59 ID O94C59 PRELIMINARY; PRT; 540 AA.  
AC O94C59;  
DT 01-DEC-2001 (TremBLrel. 19, Created)  
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TremBLrel. 23, Last annotation update)  
DE Putative phosphatidyl-inositol-transfer protein.  
GN T518.14.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosida II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;

```

[1]
RN SEQUENCE FROM N.A.
RP Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,
RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.,
RT "Full Length cDNA of gene TS18.14 (GI:4587525).";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY035162; XAKS9666.1; -.
DR InterPro; IPR001251; CRAL_TRIO.
DR Pfam; PF00650; CRAL_TRIO; 1.
DR PRINTS; PR00180; CRETINALDHP.
DR SMART; SM00516; SEC14; 1.
DR PROSITE; PS50191; CRAL_TRIO; 1.
SQ SEQUENCE 540 AA; 61189 MW; 0C2592EB18ACFAC6 CRC64;

Query Match 12.3%; Score 104; DB 10; Length 540;
Best Local Similarity 25.3%; Pred. No. 9.2;
Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

QY 1 TTVKEFILNKDTGEVSELKPHRVVTIIONGKEMSTIVSEEDFILPVYKGELEKGYQFDG 60
DB 27 TTVKA-VVEETKVEDESKP-----EGVEKSASFKEESDFPADLKESEKK----- 70
QY 61 WEISGFEKKDAGYVINISKDTFIKPVFK-----IEEKEEENKPTFDVSKKKNQPNVNS 117
DB 71 -ALSDLKSLKEAIVDN----TLLTKKKKSSPMKEKEEVVYKPAEVEKKKE--EAAEE 123
QY 118 OLNSHREKEDLQREHSQKSDSTKDVATVLDKNNISS 155
DB 124 KVEBKSEAVVTEAPKAETVEAVVTEIIPKEEVT 161

RESULT 24
O44016 PRELIMINARY; PRT; 325 AA.
AC O44016;
DT 01-JUN-1998 (TremBLrel. 06, Created)
DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE G5 ORF.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NC4;
RX MEDLINE=98198836; PubMed=9539429;
RA Rieben W.K. Jr., Gonzales C.M., Gonzales S.T., Pilkington K.J.,
RA Kiyosawa H., Hughes J.E., Welker D.L.;
RT "Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera related to
the Ddp1 and Ddp2 plasmid families.";
RL Genetics 148:1117-1125(1998).
DR EMBL; U00796; AAC18634.1; -.
SQ SEQUENCE 325 AA; 38448 MW; 69A43D0C632058A6 CRC64;

Query Match 12.2%; Score 103.5; DB 5; Length 325;
Best Local Similarity 23.8%; Pred. No. 5.7;
Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;

QY 21 HRVTVTIONGKEMSTIVSEEDFILPVYK-CEL--EKGYQFDGWEISGFEK----- 69
DB 57 HRTITSIGN--RFSVKXIGDBEKLFRISKNGELIVLNELEFDNFHFK--EGHLRKSKMF 112
QY 70 ---KDAGYV-----INLSKDTIKPV-----FKK----- 90
DB 113 NHIKSGYATNREIEIFLESTCTCKITQAKNSYKGRNIINKLPEEEEEEEEEEE 172
QY 91 --IEEKEEENKPTFDVSKKKNQPNVNSHSHRKE-----DLQREHSQKSDSTK 141
DB 173 EEESEEEVEKPTISEEEETPAVSEEEKEEEEEETPAVSEEEKEEEEEQEDKEK 232
QY 142 D 142
DB 233 D 233

RESULT 25
Q9LH98
RN SEQUENCE FROM N.A.
RP Yamada K., Chan M.M., Chang C.H., Dale J.M., Huan V.W., Lee J.M.,
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007060; AAD25756.1; -.
DR EMBL; BT000959; AAN41359.1; -.

```

```
ID Q9LH98 PRELIMINARY; PRT; 2081 AA.
AC Q9LH98;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T13N8.
DE Arabidopsis thaliana (Mouse-ear cross).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones";
RL DNA Res. 7:217-221(2000).
DR EMBL; AF002057; BAB03174.1; -.
SQ SEQUENCE 2081 AA; 232851 MW; D3603E1F85EFFF29 CRC64;

Query Match 12.1%; Score 103; DB 10; Length 2081;
Best Local Similarity 26.0%; Pred. No. 47;
Matches 46; Conservative 29; Mismatches 72; Indels 30; Gaps 7;

QY 7 ILNKDTGEVSELK---PHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEI 63
Db IVERNGKEDSIKESGSDGKTVEINGEELSTEEGSKD-----GKIEEGK--EGKEN 1688
QY 64 SFEGEKD-----AGVINLSKDTFIKPVFKKIEEKEENKP--TFDVS-KKD 110
Db STKEGSKDKIKIEGMEKENSTKESKDKINEIHGDKKATMBEGSKGGTNGTGKSKD 1748
QY 111 NPVNHSQLNESHKEDLQ---REHSQKSDTKDVTATVLDKNNISSKSTNNPN 163
Db SKSVEINGVDDSLKDDSKNGDINEINCKEDSVKDNVTIEQNDNLSNTSSSEPN 1805

RESULT 26
Q81226 PRELIMINARY; PRT; 238 AA.
AC Q81226;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN P10765W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
```

```
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929356; CAD51839.1; -.
KW Hypothetical protein.
SQ SEQUENCE 238 AA; 28481 MW; DA175A2B5A109B49 CRC64;

Query Match 12.1%; Score 102.5; DB 5; Length 238;
Best Local Similarity 28.0%; Pred. No. 4.8;
Matches 49; Conservative 26; Mismatches 65; Indels 35; Gaps 10;

QY 10 KDTGEVSELKPHRVTVTIQNGKMSSTIVSEEDFILPVYKGELEKGYQFDGWEISG---F 66
Db KEKEXTQETKHD-----ENNKSNVNVNSSENF-MKIYKNLKNQYHVEKFKKNQENIE 91
QY 67 EGKDA-----GYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKKDN--PQV--N 115
Db ESKKATSRKKIKLNFKLKNNLIEKVIKK--KAFLEAKNTVFPSLKKYNTVPHISNN 148
QY 116 H--SQLNESHKED-----LQREHSQKSDTKDVTATVLDKNNISSKSTNNPN 163
Db HNISQNDHNKKQDKNLIKLIKSNKDDNN-----NINNKNNNQKCDTKNLN 198

RESULT 27
Q9PPL5 PRELIMINARY; PRT; 312 AA.
AC Q9PPL5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Putative membrane protein.
GN CJ0692C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139076; CAB72966.1; -.
KW Complete proteome.
SQ SEQUENCE 312 AA; 37221 MW; 0004FA7836A741E8 CRC64;

Query Match 12.0%; Score 101.5; DB 16; Length 312;
Best Local Similarity 25.0%; Pred. No. 7.6;
Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;

QY 5 EFLINKDTGEVSELKPHRVTVTIQNGKMSSTI---VSEEDFILPVYK-----GELE 53
Db DIKERQNIROKWLKFSR---AINQKPLDDDLRDEISSDDILRRRFKKTNPKNFLEELD 61
QY 54 KYQFDGWEISGFEKGAGYVINI---SKDTFIKPVFKKIEEKEEENKPT---FDVSK 107
Db EYESKHTKSNLYLKED---LINVKLEKQSLAKKIFSKMKERKEENKTKQNPLFSR 118
QY 108 KDNP-----QVNHSQLNESHKEDIQREHSQKSDTKDV--TATVLDKNNISSK--STT 159
Db KANIKNIQTKYQTKSNQATTTQKQKELTNSIEKIQTETKIQKPLIEKLDVK 178
QY 160 NNPN 163
Db 179 NQPN 182
```

```

RESULT 28
Q9V7J0
ID Q9V7J0 PRELIMINARY; PRT; 382 AA.
AC Q9V7J0; Q9GQ81;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE CG8421 protein (Aspartyl beta-hydroxylase variant 2).
GN ASPH OR CG8421 OR CG18658.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Butris J.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Glodek A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Godek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
[2]
RN
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;

```

---

```

RT "Sequencing of Drosophila melanogaster genome.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradscky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RN Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
[6]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564328; PubMed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Ruber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved
RT Isoform of Asph Missing the Catalytic Domain Share Exons with
RT Junction.";
RL J. Biol. Chem. 275:39543-39554 (2000).
DR EMBL; AE003808; AAF58063.2; --
DR EMBL; AF289494; AAG40807.1; --
DR FlyBase; FBGN0034075; Asph.
SQ SEQUENCE 382 AA; 43287 MW; 60E5C03AEBFC6E8B CRC64;
Query Match 12.0%; Score 101.5; DB 5; Length 382;
Best Local Similarity 24.5%; Pred. No. 9.5;
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;
Qy 41 EDFILPVYKLEKGYQFDGW-----EISGFEKKDAGYVI-----NLSKDTIK 85
Db 78 EDLTPLESERFSK--VPDGVVDEHDEHGHVQEPSEALDDHDEHDDHDEDEE 135
Qy 86 PVFKIKIEKKKEENKPT-----FDVSKKDNPOVNSQLNESHKEDLPQREHSOKSDS 139
Db 136 PLTBEEELSEELEPEEPEEPADEYEDEENNA--GENTAEADREEEEDND 193
Qy 140 TKDVTATVLDKNISSKST 158
Db 194 EGTVEATVEATTEATTEAT 212
ID Q9V719 PRELIMINARY; PRT; 556 AA.
AC Q9V719;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG8421 protein.
GN ASPH OR CG8421 OR CG18658.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

```

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svatek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Waesman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*,"  
RL Science 287:2185-2195(2000).  
RN [2]  
SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banjon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome,"  
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
SEQUENCE FROM N.A.  
RA Mixra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergsma C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome,"  
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.  
DR

DR FlyBase: FBgn0034075; Asph.  
SQ SEQUENCE 556 AA; 63144 MW; B420980CDB6C357A CRC64;  
Query Match 12.0%; Score 101.5; DB 5; Length 556;  
Best Local Similarity 24.5%; Pred. No. 14;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
QY 41 EHFILPVVKGLEKGYQFDGW-----HISGFEKKDAGYVI-----NLKDTFTK 85  
DB 78 EDLDTPLSRSFSK--VFDGWVDEHDDHGDVQPSGEALDDHDDHDDHDEDEDE 135  
QY 86 PVFKIEEKKEENKPT-----FDVSKKKNQPNVHSHLNESHKRELQREHSOKS 139  
DB 136 PITELELELEEEFEFTDEPAADVEEYDEDEENNA--GENIATADEAEEREDND 193  
QY 140 TKDVTATVLDKNNISKST 158  
DB 194 EGTVEATVEATTEATTEAT 212  
RESULT 30  
Q9GQ82  
ID Q9GQ82 PRELIMINARY; PRT; 785 AA.  
AC Q9GQ82;  
DT 01-MAR-2001 (TREMELrel. 16, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DE 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA).  
GN ASPH OR CG8421 OR CG18658.  
OS *Drosophila melanogaster* (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; *Drosophila*.  
ON NCBI Taxid=7227;  
RX [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20564328; PubMed=10956665;  
RA Dinchuk J.B., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
RA Friedman P.A.;  
RT "Aspartyl beta-hydroxylase (Asph) and an evolutionarily conserved  
RT isoform of Asph missing the catalytic domain share exons with  
RT Junction.";  
RL J. Biol. Chem. 275:39543-39554(2000).  
RN [2]  
SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svatek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Waesman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*,"  
RL Science 287:2185-2195(2000).  
RN [2]  
SEQUENCE FROM N.A.  
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banjon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome,"  
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
SEQUENCE FROM N.A.  
RA Mixra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Bergsma C., Berman B., Carlson J.W., Celniker S.E.,  
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome,"  
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RN Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.  
DR



RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kimos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
[3]  
RP SEQUENCE FROM N.A.  
RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Anantides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Buzam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[4]  
RP SEQUENCE FROM N.A.  
RA Mirza S., Crebby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Berniker B., Carlson J.W., Celnik S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[5]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[6]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF289493; AAG40806.1; -  
DR EMBL; AE003808; AAM70947.1; -  
DR FlyBase; FBgn0034075; Asph.  
DR InterPro; IPR001440; TPR.  
DR InterPro; IPR006025; Zn MTpeptidse.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
SQ SEQUENCE 785 AA; 89843 MW; 30A8DFCD6836F7F1 CRC64;  
Query Match 12.0%; Score 101.5; DB 5; Length 785;  
Best Local Similarity 24.5%; Pred. No. 21;  
Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
QY 41 EDFILPVYKGEKCYQFDGW-----EISPEGKADAGYVI-----NLSKDTFIK 85  
DB 78 EDLDTPLSESRFSK--VFQGVDEHRDHDGHDVQEPSEALDDHDEHDDHDDHDEDEE 135  
QY 86 PVFKIEKKEENKPT-----FDVSKKKQNPQVNSQLNESHKEDLQREHSQKSDS 139  
DB 136 PLTESLEEELEEEPTDEDEPAADBEYEDEDEENNA--GENITAEADAESEEDND 193  
QY 140 TKDVTATLTKNNISKST 158  
DB 194 EGTVEATVEATTEAT 212

RESULT 31  
Q9U4U6 PRELIMINARY; PRT; 948 AA.  
AC Q9U4U6;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Ornithine decarboxylase.  
GN ODC  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FCUP1/RSA;  
RA Birkholtz L., Joubert F., Neitz A.W.H., Louw A.I.;  
RT "Molecular characterisation of plasmodium falciparum ornithine  
RT decarboxylase cDNA obtained by RACE.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF139900; AAF14518.1; -  
DR HSSP; P07805; 1F3T.  
DR InterPro; IPR000183; Decarboxylase2.  
DR Pfam; PF02784; Orn Arg dec N; 1.  
DR Pfam; PF02078; Orn DAP Arg dec; 1.  
DR PRINTS; PR01179; ODADCRXLASE.  
SQ SEQUENCE 948 AA; 110350 MW; 43F103DB83F12835 CRC64;  
Query Match 12.0%; Score 101.5; DB 5; Length 948;  
Best Local Similarity 21.8%; Pred. No. 26;  
Matches 51; Conservative 31; Mismatches 63; Indels 89; Gaps 11;  
QY 6 FILAKDNGEVEL-----KPHRVTVTIQNGKMSSTIVSEEDFLPVY-----KG 50  
DB 559 YIIINLGGYPEELDYNAKGDKIHYCTLSLQEIKKDIQKFLNEETFKTKGYYSFKI 618  
QY 51 ELEKGYQFDGWEISGFEKGD-----AGYVINLSKDTFIKPVFKIEKKEENKPT 102  
DB 619 SLAINMSIDHY-----FSHMKDNLRAVCEPGRYVAASSTLAVKIIGK-----RPT 665  
QY 103 F-----DVSKKQNPQVNSQLNESHKED-----LQREHSQK----- 136  
DB 666 FQGIHLKDLKAHYDPLNFAQOENKQDEPKINHNNNDNNNDNNNNNNNNNNKQGG 725  
QY 137 -----SDST-----KDYATATVLDK--NNIS-SKSTTNPN 163  
DB 726 GNIMNDLIITNDSTNKNKNDHSSSQVQNVSCTIRKEGDKINHTHTINPN 779  
RESULT 32  
Q81556 PRELIMINARY; PRT; 1377 AA.  
AC Q81556;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Eukaryotic translation initiation factor 3 subunit 10, putative.  
GN PFL0625C.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3D7;  
RX MEDLINE=22255705; PubMed=12368864;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,  
RA Perteira M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,



```
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RL Nature 419:498-511(2002).
RL EMBL; AB014846; AAN36214.1; -.
KW Initiation factor.
SQ SEQUENCE 1377 AA; 166059 MW; 587CF2E3F2C8FB99 CRC64;

Query Match 12.0%; Score 101.5; DB 5; Length 1377;
Best Local Similarity 22.3%; Pred. No. 39;
Matches 43; Conservative 37; Mismatches 62; Indels 51; Gaps 9;

QY 3 VKEFLNKDTGSEVSELPKPRVTVTTQNGKE--MSSTIVSEEDFILPVYKGELEKGVQF-- 58
DB 619 IKDSLINMYNDLQGLKMNITAIENEVENITMSTLTYYEEBF--AYLDKSNKNINVT 675
QY 59 ----DGEWISGPEGKDDAGVINLSKDTFKPVFKKIEK-----KEEENKPRFVSK 107
DB 676 ATLEDIIEVEFEK-----IENEKLLKIIYEKIDDEHHKIQLLSEHNKKRKLK 728
QY 108 KXDNPOVNHSQLN-----ESHKEDLQR-----EHSQKSDSTKDVVAT 146
DB 729 KQ--KELEQAQLKMKELKLEKKEBLARKGKELRIEKKHKKTEAAEQMLKE 786
QY 147 VLDKNNISSKSTT 159
DB 787 I---KKLCSTNTT 796

RESULT 33
Q81353 PRELIMINARY; PRT; 1373 AA.
AC Q81353;
DT 01-WAR-2003 (TremBLrel. 23, Created)
DT 01-WAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-WAR-2003 (TremBLrel. 23, Last annotation update)
DE Helicase with Zn-finger motif, putative.
GN PFI0480W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Lake N., Lawson D., Lennard N.,
RA Line A., Maddison M., Melean J., Mooney P., Moulé S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrell B.G.
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR EMBL; AL929356; CAD51782.1; -.
SQ SEQUENCE 1373 AA; 160372 MW; 059FB4E99EEF4702 CRC64;

Query Match 11.9%; Score 100.5; DB 5; Length 1373;
Best Local Similarity 21.7%; Pred. No. 45;
Matches 40; Conservative 33; Mismatches 54; Indels 57; Gaps 9;

QY 6 FILNKDTGVSE----LKPRHRTVTIQNGKMSSTIVSEEDFILPVYKGELEKGVQFGW 61
DB 418 YVNDGRGVWEESIIMLPHHQILL-----LSATV-----PNY-----LEFADW 457
QY 62 -----EISGPEGKK-----DAGVINLSKDTFKPVFK-----KIEKKERE 98

RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RL Nature 419:498-511(2002).
RL EMBL; AB014846; AAN36214.1; -.
KW Initiation factor.
SQ SEQUENCE 1377 AA; 166059 MW; 587CF2E3F2C8FB99 CRC64;

Query Match 12.0%; Score 101.5; DB 5; Length 1377;
Best Local Similarity 22.3%; Pred. No. 39;
Matches 43; Conservative 37; Mismatches 62; Indels 51; Gaps 9;

QY 3 VKEFLNKDTGSEVSELPKPRVTVTTQNGKE--MSSTIVSEEDFILPVYKGELEKGVQF-- 58
DB 619 IKDSLINMYNDLQGLKMNITAIENEVENITMSTLTYYEEBF--AYLDKSNKNINVT 675
QY 59 ----DGEWISGPEGKDDAGVINLSKDTFKPVFKKIEK-----KEEENKPRFVSK 107
DB 676 ATLEDIIEVEFEK-----IENEKLLKIIYEKIDDEHHKIQLLSEHNKKRKLK 728
QY 108 KXDNPOVNHSQLN-----ESHKEDLQR-----EHSQKSDSTKDVVAT 146
DB 729 KQ--KELEQAQLKMKELKLEKKEBLARKGKELRIEKKHKKTEAAEQMLKE 786
QY 147 VLDKNNISSKSTT 159
DB 787 I---KKLCSTNTT 796

RESULT 34
Q14712 PRELIMINARY; PRT; 375 AA.
AC Q14712;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Cell cycle progression restoration 8 protein.
GN CPR8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98043401; PubMed=9383053;
RA Edwards M.C., Liegeois N., Horecka J., DePinho R.A., Sprague G.F. Jr.,
RA Tyers M., Elledge S.J.;
RT "Human CPR (cell cycle progression restoration) genes impart a Far-
RT phenotype on yeast cells."
RL Genetics 147:1063-1076(1997).
DR EMBL; AF011794; AAB69314.1; -.
DR InterPro; IPR004238; LEA.
DR Pfam; PF02987; LEA; 1.
SQ SEQUENCE 375 AA; 44320 MW; 0C261BF68AAEF7AD CRC64;

Query Match 11.8%; Score 100; DB 4; Length 375;
Best Local Similarity 28.0%; Pred. No. 12;
Matches 51; Conservative 16; Mismatches 67; Indels 48; Gaps 9;

QY 13 GEVSELPKPRVTVTTIQNGKMSSTIVSE--EDFILPVYKGELEKGVQF-----DGWE--- 62
DB 18 GELQQLSGSQL-----HGKSDSPNVTKEKAIILRLTELEKLTTEQQRSDLWRLY 72
QY 63 -----ISGPEGKDDAG-----YVINLSKDTFKPVFKKIEKKEENKPTFDVSKKKD 110
DB 73 VEAKDQNGKQGTGDKKGGRGSHRVKNSKGTFLGSV-----KETFDAMKNST 120
QY 111 NPQVNHSQLNESHKEDIQREHSQKSDST-----KDVATVLD-KNNISSKSTTNP 162
DB 121 KEFVRRHKEIKQAKEDV-KENLKFSVSVKSTFRHFDTTKNIFDEKGNKFNATKAA 179
QY 163 NK 164
DB 180 EK 181

RESULT 35
Q890A1 PRELIMINARY; PRT; 622 AA.
AC Q890A1;
DT 01-OCT-2001 (TremBLrel. 18, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-OCT-2002 (TremBLrel. 22, Last annotation update)
DE LIPOPROTEIN.
GN MYP_4650.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2107;
RN [1]
RP SEQUENCE FROM N.A.
```



Search completed: February 10, 2004, 10:57:09  
Job time : 31.0366 secs





Db 218 KVENKSGDHIKEVKKVEHEDEBE-----DKKEKSKNKNDK 261

## RESULT 6

D86432

hypochemical protein T518.14 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001

C:Accession: D86432

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: D86432

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-540 &lt;STO&gt;

A:Cross-references: GB:AE005172; NID:g4587525; PIDN:AAD25756.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 12.3%; Score 104; DB 2; Length 540;

Best Local Similarity 25.3%; Pred. No. 2.8;

Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

Qy 1 TTVEFLNKDTGSEVSELPKRVVTIONGKEMSSSTVSEDFILPVKGELEKGYQEDG 60

Db 27 TTVKA-VVETKVEDESKP-----EGVEKSAFKEESDPFADLKESEK----- 70

Qy 61 WEISFEGKQAGYINLSKOTFIKPVFK--IEKKEENKPTFDYKSKDNQVNH 117

Db 71 -ALSLKSLTEAIVDN---TLTKKKESSPMKEEVEKPEAEVKKKE--EAAEE 123

Qy 118 QLNESHKEDLQREHSQKSDSTKDTATVLDKNISS 155

Db 124 KYEEKSEAVVTEAPKATVAVTEIIPKEVTT 161

## RESULT 7

T18283

hypochemical protein G5 - slime mold (Dictyostellium discoideum)

C:Species: Dictyostellium discoideum

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T18283

R:Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington, K.J.; Kiyosawa, H.; Hugh

Genetics 148, 1117-1125, 1998

A:Title: Dictyostellium discoideum nuclear plasmid Dd65 is a chimera related to the Dd61

A:Reference number: Z14684; MUID:98198836; PMID:9539429

A:Accession: T18283

A:Status: Preliminary;

A:Molecule type: DNA

A:Residues: 1-325 &lt;RIE&gt;

A:Cross-references: EMBL:U00796; NID:g2702254; PID:g2702258; PIDN:AAC18614.1

C:Genetics:

A:introns: 85/1

Query Match 12.2%; Score 103.5; DB 2; Length 325;

Best Local Similarity 23.8%; Pred. No. 1.7;

Matches 43; Conservative 25; Mismatches 50; Indels 63; Gaps 9;

Qy 21 HRTVTTIONGKEMSSSTVSEDFILPVK-GEL--EKGYQFGWEISGFEK----- 69

Db 57 HRTVTSIKK--RFSVYKIDDEKLFRISSKGLIYVNLAEFDNFHKK--EGHRLKRSKWF 112

Qy 70 ---KQAGV-----INSKDTFIKPV-----FKK----- 90

Db 113 NHIKSGYVATNEIEIFLESECTLCKETTAQTKNSYKGRNIINKLPREEEEEEEE 172

Qy 91 --IEKKEENKPTFDVSKKDNQVNHSLNESHRK-----DLQREHSQKSDSTK 141

Db 173 EEEQEEVEKPTISEEEBETPAVSEKEEEEEEEETPAVSEKEEEEEQEDKEX 232

Qy 142 D 142

Db 233 D 233

## RESULT 8

G81339

probable membrane protein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 11169)

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C:Accession: G81339

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Baaham, D.; Chilling-

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrel

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: G81339

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-312 &lt;PAR&gt;

A:Cross-references: GB:AL139076; GB:AL111168; NID:g6968128; PIDN:CAB72966.1; PID:g696814

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj0692c

Query Match 12.0%; Score 101.5; DB 2; Length 312;

Best Local Similarity 25.0%; Pred. No. 2.3;

Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;

Qy 5 EFILNKDTGSEVSELPKRVVTIONGKEMSSSTI---VEEDFILPVK-----GELE 53

Db 5 DPKERQNIROKMLKFSR--AINQKPLDDLRPEISSDILRRPKKTKPNKLEELD 61

Qy 54 KGYQFGWEISGFEKQAGYVNI---SKDTFIKPVFKLIEKKEENKPT--FDVSK 107

Db 62 EYESKHTKKSNIYKED---LINVKLEKQSLAKKIFSKMKERKKEKTKKQNFLE 118

Qy 108 KKNP---QVNHSLNESHRKEDLQREHSQKSDSTDV--TATVLDKNISSK--STT 159

Db 119 KKAIEIKNIQTQTQTSNOATTQTKQEKELTMSIEKIKTEKIKQPIIEKKLPV 178

## RESULT 9

A90570

lipoprotein [imported] - Mycoplasma pulmonis (strain VAB CT1P)

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001

C:Accession: A90570

R:Chamblaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;

Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm

A:Reference number: A9512; MUID:21267165; PMID:11153084

A:Accession: A90570

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-632 &lt;KUN&gt;

A:Cross-references: GB:AL445566; PID:g14089879; PIDN:CAC13638.1; GSPDB:GN00153

A:Experimental source: strain VAB CT1P

C:Genetics:

A:Gene: MYPV 4650

A:Genetic code: SGC3

Query Match 11.8%; Score 100; DB 2; Length 622;

	Best Local Similarity	24.4%	Pred. No. 6.5;	Indels	38;	Gaps	77
	Matches	40;	Conservative	31;	Mismatches	55;	
Oy	8	LNKDTGESELSKPHRVTVTIIQNGKMSSTIYSEEDFILPYVKGLELGYDFDGEWISGFE	67				
Db	69	INSETFELVTKVE-----TIIN-----LSEDIIITSLNTEENAKFOLDLF-----	108				
Oy	68	GKKDAGYIINSKDFTIKRVPFKLTIEEKEENKPFVDYSKKD-----NPVVNSHOLNES-	122				
Db	109	-----VSKEKFKEIKFOELINFSGTEQ-KITDNTSSDEDKKNPKPKNENSNNNS	157				
Oy	123	-HRKEDLQREBSHQSDSKDVTAIVLLKN-NISKSTTTNNPNK	164				
Db	158	DQKDELQKNNSDKLNDNVQDEKAKENSNDSSEKQDENVTNK	201				

	RESULT	10
T20410	hypothetical protein E02A10.2 - Caenorhabditis elegans	
C/Species:	Caenorhabditis elegans	
C/Date:	15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999	
R/Accession:	T20410	
R/Thomas, K.	submitted to the EMBL Data Library, October 1996	
A/Reference number:	Z19271	
A/Accession:	T20410	
A/Status:	Preliminary; translated from GB/EMBL/DDBJ	
A/Molecule type:	DNA	
A/Residues:	1-385 <MWt>	
A/Cross-references:	EMBL:Z81053; PIDB:CAB02877.1; GSPDB:GN00023; CESP:E02A10.2	
A/Experimental source:	clone E02A10	
A/Genetics:		
A/Gene:	CESP:E02A10.2	
A/Map position:	5	
A/Intons:	32/1; 72/1; 85/1; 122/1; 133/1; 220/3	
Query Match:	11.7%; Score 99.5; DB 2; Length 385;	
Best Local Similarity:	27.7%; Pred. NO.4.1'	
Matches	46; Conservative 26; Mismatches 51; Indels 43; Gaps 8;	
Oy	2 TVEKEFILNKD-----TGEVSELEKPHRVTVITQ--NGKMSSTIYSEDPFI 44	
Db	232 TIKEEVAQAQDIYIVCNEKPAPFTATETDFCSGLQKNVACTILRIINHKVAKR--NEEDKK 289	
Oy	45 LPVYKGSELGKYQFDMWEISGFEGKDGAGYVNLNKTQFIKPVFKKIEEKKEENKPTFD 104	
Db	290 EEPRKGEBEKKGEVYKKE--EDEKGD-----EP--KKEBEKKEEBKKE--E 330	
Oy	105 VSKKNDPQVNHSQLNSHREKDLQREHSQKSQSDTKDVATVTLDK 150	
Db	331 VEKKEE-----EEKDEPRKKEBEKKEEBEKKELVEKESKEVKEK 371	

```

RESULT 11
T25911
hypothetical protein T23B3.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25911
C:Maggi, L.; Le, T.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid T23B3.
A:Reference number: Z20109
A:Accession: T25911
A:Status: preliminary; translated from GB/EMBL/DD5J
A:Molecule type: DNA
A:Residues: 1-211 <MAG>
A:Cross-references: EMBL:U086309; PIDD:AA042334.1; GSPDB:GN000019; CESP:T23B3.5
A:Experimental source: Strain Bristol N2; clone T23B3
C:Genetics:
A:Gene: CESP:T23B3.5
A:Map position: 1
A:introns: 30/2; 200/3

```

[illegible]

RESULT 12  
 B72291  
 hypochlorite protein - *Thermotoga maritima* (strain MSB8)  
 C|Species: *Thermotoga maritima*  
 C|Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
 C|Accession: B72291  
 C|Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
 C.M.  
 Nature 399, 323-329, 1999  
 A|Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
 A|Reference number: A72200; MUID:99287316; PMID:10360571  
 A|Accession: B72291  
 A|Status: preliminary  
 A|Molecule type: DNA  
 A|Residues: 1-219 <ARN>  
 A|Cross-References: GB:AE001771; GB:AE000512; NID:g49981678; PIDN:ADM36218.1; PID:g4998168  
 A|Experimental source: strain MSB8  
 A|Genetics:  
 A|Gene: TM1142

Query Match	Similarity	11.7%	Score 99	DB 2	Length 219
Best Local	26.4%		Pred. No. 2.4		
Matches	46	Conservative	30	Indels	42
				Gaps	10
Qy	4	KEFLINDQGEVSLKPRVTVTIQNMKEMSSITVSEEDFILPYVKGELKGYQ--PDGW	61		
Db	52	KEVFPKEDPQEI--LIINFVPELVYXDEKTLTFAK-----VLGEESVSYRDLPLAGF	102		
Qy	62	EISG-----FEGRKDAGYVIN-LSKDTFLKPVFKLIEKKEENKFTPDVSKKKNPQV	114		
Db	103	GVRSPTPTFFPKGKGGGLPYGVYDKNFIK-ILKYVAQELKED-----FQTYLKKDDPFV	157		
Qy	115	NHSQLNASHRED-----LQREHSQKSDS-----TKQVATATVLDK	150		
Db	158	GEPLILIF-KEADFVLEKDEKNAVXQDTVANEVRDRIRIYVTSPPDAKTLQEK	210		

RESULT 13  
 AB4152  
 hypothetical protein BH4017 [imported] - *Bacillus halodurans* (strain C-125)  
 C|Species: *Bacillus halodurans*  
 C|Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C|Accession: AB4152  
 R|Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A|Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
 A|Reference number: AB6650; MUID:20512582; PMID:11058132  
 A|Accession: AB4152  
 A|Status: preliminary  
 A|Molecule type: DNA  
 A|Residues: 1-614 <STO>  
 A|Cross-references: GB:AP001520; GB:BA000004; NID:gl0176401; PIDN:BA007736.1; GSPDB:GN00  
 C|Experimental source: strain C-125  
 C|Genetics:  
 A|Gene: BH4017

Query Match	11.5%	Score 97.5	DB 2	Length 614
Best Local Similarity	27.3%	Pred. No. 9.8		



Matches 41; Conservative 26; Mismatches 44; Indels 39; Gaps 8;

QY 16 SELKHHRTVTTQNGKEMSTVSE-----EDFILPVYKGELEKGYQFDG---EISG 65  
DB 386 SEPKEE--TYTLQTAIQM--TPVINEYSPQTRSEFL-----ARKAHQIDGMADVSKVY 435  
QY 66 FEGKDGAVINILSKDTFKPVFKIEKK--EENKPTFDVSKKK---DNPQVNHSQL 119  
DB 436 FEGKINIASQL-----GKIEBKLDQKINNTFFDFPKKEVNVQSEIKSTSL 482  
QY 120 NESHREKDLQREHSGKSDSTKDVATVLD 149  
DB 483 GKVIIGIVDGRKRYHKGKEDLERLSKQIE 512

## RESULT 14

S46817  
hypothetical protein YHR080C - Yeast (Saccharomyces cerevisiae)  
C.Species: Saccharomyces cerevisiae  
C.Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 19-Apr-2002  
C.Accession: S46817  
R.Favell, T.  
submitted to the EMBL Data Library, June 1994  
A.Description: The sequence of S. cerevisiae cosmid 9205.  
A.Reference number: S46795  
A.Accession: S46817  
A.Molecule type: DNA  
A.Residues: 1-1145 <FAV>  
A.Cross-references: EMBL:U10556; NID:g500825; PIDN:AAB68895.1; PID:g500838; MIPS:YHR080C  
C.Genetics:  
A.Map position: 8R  
A.Cross-references: SGD:S0001122  
C.Keywords: transmembrane protein

Query Match 11.5%; Score 97.5; DB 2; Length 1345;  
Best Local Similarity 27.0%; Pred. No. 24;  
Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

QY 49 KCELEKGYQFDGMEISGFEGK--DAGYVILNSKDTFKPVFKIEKKEENKPTFDVSK 107  
DB 1109 KCAIEKG-----SVEGKQVSDVYMLSELDI-----SRASKKPVKQVVK 1149  
QY 108 KKDNPQVNHSQLNESHREKDLQREHSGKSDSTKDVATVLD--KNNISKSTN---NP 162  
DB 1150 SHDKRRPFRFSKVE-----QKSSRSKSDNDKILTHLDPVQNNFSSSEIFMNNKLSP 1201  
QY 163 NK 164  
DB 1202 QK 1203

## RESULT 15

T14188  
hypothetical protein T28D5.30 - Arabidopsis thaliana  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 28-Jul-2000  
C.Accession: T14188  
R.Bevan, M.; Lemard, N.; Quail, M.; Harrie, B.; Rajandream, M.A.; Barrell, B.G.; Bancro  
submitted to the Protein Sequence Database, August 1999  
A.Reference number: Z17931  
A.Accession: T14188  
A.Molecule type: DNA  
A.Residues: 1-988 <BEV>  
A.Cross-references: EMBL:AL109819  
A.Experimental source: cultivar Columbia; BAC clone T28D5  
C.Genetics:  
A.Gene: ATSP-T28D5.30  
A.Map position: 4  
A.Introns: 162/3; 201/3; 416/3; 438/3; 460/3; 482/3; 504/3; 519/3; 534/3; 559/3; 579/3;  
C.Superfamily: Arabidopsis thaliana hypothetical protein T2IC14.40

Query Match 11.4%; Score 97; DB 2; Length 988;  
Best Local Similarity 22.1%; Pred. No. 18;

Matches 34; Conservative 31; Mismatches 67; Indels 22; Gaps 5;

QY 29 NGEKMSSTIVSEEDITLPVYKGELEKGY-----QFDGMEISGFEGKDGAVINILSKDTF 83  
DB 383 NGRQRNSNVQSSVDILSYTIDKVPGLNVSEHDIVELVDVRSAGSLSPVQORD- 441  
QY 84 IKPVFKIEEKK-----EENKPTFDVSKKKDNPQVNHSQLNESHREKDLQRE 132  
DB 442 VEPVGDVRSQSDMSPNPSAANNVREGPATPFIMSEDDPGNDVAPMEDHIRSEVQLSP 501  
QY 133 HSQKSDSTKDVLT--ATVLDKNNISKSTNNPNK 164  
DB 502 HVL---GAKDVTVDSDPTDKVGVNDVTDAADPTE 532

## RESULT 16

T05612  
hypothetical protein F9D16.270 - Arabidopsis thaliana  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
C.Accession: T05612  
R.Bevan, M.; Medler, H.; Medler, E.; Wambutt, R.; Hohnsels, J.; Mewes, H.W.; Meyer, K.F.  
submitted to the Protein Sequence Database, February 1999  
A.Reference number: Z15419  
A.Accession: T05612  
A.Molecule type: DNA  
A.Residues: 1-456 <BEV>  
A.Cross-references: EMBL:AL035394  
A.Experimental source: cultivar Columbia; BAC clone F9D16  
C.Genetics:  
A.Map position: 4  
A.Introns: 110/3; 247/2; 282/3; 304/3; 361/3; 390/2; 418/3  
A.Note: F9D16.270

Query Match 11.4%; Score 96.5; DB 2; Length 456;  
Best Local Similarity 25.0%; Pred. No. 8.3;  
Matches 40; Conservative 29; Mismatches 60; Indels 31; Gaps 8;

QY 10 KDTGEVSELKP--HRVTITQNGKEMSTVSEEDITLPVYKGELEKGYQFDGMEISGFEG 68  
DB 245 KKKKEKDKLPKGPVSAFLVYANERRAALREENKSVVEVAK-----ITGEWKNLSD 296  
QY 69 KKDAGY--VINLSKDTFKPVFKIEKKEENKPTFDVSKKKDNPQVNHSQLNESHRE 126  
DB 297 KKAQYERKAKKKNKETYIQ--AMEEYKRTKEE-----ALSQKE-----EEELKLDKQE 345  
QY 127 DLQREHSGKSDSTKDVATVLDKNNISKSTNN--PNK 164  
DB 346 ALQMLKKKEKTDN-----LIKKEKATKKKKNEVNDENK 378

## RESULT 17

T37189  
hypothetical protein C02H7.1 - Caenorhabditis elegans  
C.Species: Caenorhabditis elegans  
C.Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000  
C.Accession: T37189  
R.Lembaac, D.; Minx, M.  
submitted to the EMBL Data Library, February 1996  
A.Description: The sequence of C. elegans cosmid C02H7.  
A.Reference number: Z20523  
A.Accession: T37189  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-535 <LEI>  
A.Cross-references: EMBL:U49945; PIDN:AA047924.1; GSPDB:GN00029; CESP:C02H7.1  
A.Experimental source: strain Bristol N2; clone C02H7  
C.Genetics:  
A.Gene: CESP:C02H7.1  
A.Map position: X  
A.Introns: 47/3; 100/3; 149/3; 304/2; 347/3; 458/3

Query Match 11.3%; Score 96; DB 2; Length 535;

Best Local Similarity 21.3%; Pred. No. 11;  
Matches 35; Conservative 28; Mismatches 61; Indels 40; Gaps 5;

QY 11 DTGEVELKPHRTVITIQNGKMSSTIVSEDFILVYKGELEKGYQFGMEISGEGKK 70  
DB 77 DQSLKNVAAKII---SGKDAET-----NKMQLMGSTNATSNNSKN 116  
QY 71 DAGVYNLSKDTFIPVFKIEKEKEENKPTFVSKKQDPQVNHSQLNESHREKEDLOR 130  
DB 117 GTC-----EKKKKKKVKKEDKKGPDEEKT---TKGSSKKEETHEKEKSKKSSABE 166  
QY 131 EEHSQKSDSTK-----DVTATVLDKNNISSKSTNNPNK 164  
DB 167 KEKKKKSSSKERHKSSDRSEKSEKSKKKEKSTTDEKPK 210

## RESULT 18

G71609

Hypothetical protein PFB0650w - malaria parasite (Plasmodium falciparum)

C/Species: Plasmodium falciparum

C/Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000

C/Accession: G71609

R/Gardner, M.J.; Tetteelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;

.; Petrea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Science 282, 1126-1132, 1998

A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A/Reference number: A71600; PMID:99021743; PMID:9804551

A/Accession: G71609

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-2500 &lt;GAR&gt;

A/Cross-references: GB:AE001408; GB:AE001362; NID:G3845238; PIDN:AAC71919.1; PID:G384524

A/Experimental source: clone 3D7

A/Genetics:

A/Gene: PFB0650w

Query Match

Best Local Similarity 11.3%; Score 96; DB 2; Length 2500;

Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;

QY 19 KPHRTVITIQNGKMSSTIVSEDFILVYKGELEKGYQFGMEISGEGKKDAGVYNL 78  
DB 2173 KPYKIT---ENNKK-----NEGNEILKYSIENEENKNNYDKQENECILDKDTQCNVMT 2223  
QY 79 -----SKDTFLKPYPKKIEEKKKEENKPTFVSKKQDPQVNHSQLN-----SHRKED 127  
DB 2224 KKKNNLDNKKSPSPNITKVKLEEEKSDDKRD---DKKNDTRKNNLDNKKSPSPNITKVK 2280  
QY 128 LQREHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164  
DB 2281 LEEEE---KSDDKRD-----DKKNDTRKNNLDNKK 2308

## RESULT 19

T47835

Hypothetical protein T209.90 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C/Accession: T47835

R/Nyakatura, G.; Fartmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;

submitted to the Protein Sequence Database, February 2000

A/Reference number: Z24475

A/Accession: T47835

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-644 &lt;NYA&gt;

A/Cross-references: EMBL:AL138658

A/Experimental source: cultivar Columbia; BAC clone T209

C/Genetics:

A/Map position: 3

A/Intons: 158/2; 329/3

A/Note: T209.90

Query Match 11.3%; Score 95.5; DB 2; Length 644;  
Best Local Similarity 23.4%; Pred. No. 14;  
Matches 34; Conservative 22; Mismatches 58; Indels 31; Gaps 4;

QY 28 QNGKMSSTIVSE---DFILPVYKGELEKGYQFGMEISGEGKKDAGVYNLSKDTF 83  
DB 519 ENSKTEKKTADVADKKSVADFLRKIKNSPQKK---ETTSKQKQNDGNV----- 565  
QY 84 IKPVFKIEKEKEENKPTFVSKKQDPQVNHSQLNESHREKEDLQREHSQKSDS--- 139  
DB 566 -----KKENDHQKSDGNVKKNSKVPRELRSSTGKKKYEVENNNSKSSKKRKQ 615  
QY 140 TKDVATVLDKNNISSKSTNNPNK 164  
DB 616 TKETAEVATGKRGRESGKDKQPRK 640

## RESULT 20

I37271

Cytich II - human

C/Species: Homo sapiens (man)

C/Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 21-Jul-2000

C/Accession: I37271; S52774

R/Bess, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.

Exp. Cell Res. 218, 174-182, 1995

A/Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: chr

A/Reference number: I37271; PMID:95255491; PMID:7737358

A/Accession: I37271

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-348 &lt;HES&gt;

A/Cross-references: EMBL:Z46788; NID:G758586; PIDN:CAA86752.1; PID:G758587

Query Match

Best Local Similarity 11.2%; Score 95; DB 2; Length 348;

Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

QY 49 KSELEKGYQFGMEISGEGKKDAGVYNLSKDTFIPVFKKIEKEKEENKPTF---DV 105  
DB 205 ESEGEKG-----GTEKDSKKGKKDS---KKGDSALIELQVADKEDKDEGKKDANKGDE 256  
QY 106 SK--KKDNPQVNHSQLN-----BSHRKEDLQREHSQKSDSTKD--VTATVLDKNNI 153  
DB 257 SKDAKKDAKEIKKGGKKKPPSTSDSKDVKKE---SKDATTDAKKVAKKDKTEKESA 313  
QY 154 SSK 156  
DB 314 DSK 316

## RESULT 21

T10466

DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) II - malaria parasite (Plasmodium falc

C/Species: Plasmodium falciparum

C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 20-Jun-2000

C/Accession: T10466

R/Cheesman, S.J.

submitted to the EMBL Data Library, September 1995

A/Reference number: Z17031

A/Accession: T10466

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1397 &lt;CHE&gt;

A/Cross-references: EMBL:K79345; NID:G994807

C/Genetics:

A/Gene: TopoII

A/Map position: 14

A/Supfamily: eukaryotic type II DNA topoisomerase; phase T4 DNA topoisomerase (ATP-hy

C/Keywords: ATP; DNA binding; isomerase; nucleus

Query Match 11.2%; Score 95; DB 2; Length 1397;

Best Local Similarity 23.8%; Pred. No. 38;

Matches 36; Conservative 36; Mismatches 53; Indels 26; Gaps 6;

Query Match	11.1%;	Score 94.5;	DB 2;	Length 871;
Best Local Similarity	25.3%;	Pred. No. 24;		

Query Match	11.1%;	Score 94;	DB 2;	Length 210;
Best Local Similarity	28.6%;	Pred. No. 5.2;		



Qy 152 N-----ISSKSTNNPNK 164  
 Db 559 SKIAESSLQEIISASQIENSPTQ 582

## RESULT 29

ORMSPI

microtubule-associated protein MAP1B - mouse

N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 01-Sep-2000

C:Accession: S07549; S44387; A33645

R:Noble, M.; Lewis, S.A.; Cowan, N.J.

J. Cell Biol. 109, 3367-3376, 1989

A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contains

A:Reference number: A33645; MUID:90094539; PMID:2480963

A:Accession: S07549

A:Molecule type: mRNA

A:Residues: 1-2464 &lt;NOB&gt;

A:Cross-references: EMBL:X51396; NID:g52999; PIDN:CAA35761.1; PID:g53000

R:Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.

Arch. Biochem. Biophys. 310, 428-432, 1994

A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.

A:Reference number: S44387; MUID:94234720; PMID:8179328

A:Accession: S44387

A&gt;Status: preliminary

A:Molecule type: protein

A:Residues: 653-663, 'IC' &lt;SAN&gt;

C:Superfamily: microtubule-binding; phosphoprotein; tandem repeat

C:Keywords: microtubule binding; phosphoprotein; status experimental &lt;EMB&gt;

F:589-592,639-642,649-652,655-656,660-663,668-671,674-677,679-682,683-686,687-690,691-69

R-K-E/D-X)

F:1861-2064/Region: 17-residue repeat

F:91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: PH

F:147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (CD

F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 11.0%; Score 93; DB 1; Length 2464;

Best Local Similarity 27.8%; Pred. No. 16+02;

Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;

Qy 4 KEFLINKDGEVSELKPHRVTVITIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGMFI 63

Db 584 EKVLYKDKDPVTKESK---SVT---EKVVSS---KEQ---SPV---KAQVA-----EK 623

Qy 64 SGFEGKQAGYVINSKDTFIRPVFK-KIEKKKEENKPTPDVSKKQDPVNHSSOLNS 122

Db 624 QATESK-----KTKDKVKKKEIKTKLEKKKE--KPKKEVKKEDKTPL---KKDK 672

Qy 123 HKEDLQRE-----EHSQKSDSTKDVATVLDKNNISSKSTNNPNK 164

Db 673 PKKEVKKKEIKKEIKKEKKEKELKEVKKETPLKDAKKEVKKKEKKEKPKK 728

## RESULT 30

T39009

hypothetical protein SPAC6B12.02c - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 21-Jan-2000

C:Accession: T39009

R:Gentles, S.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, August 1995

A:Reference number: Z21815

A:Accession: T39009

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-188 &lt;GEN&gt;

A:Cross-references: EMBL:298531; PIDN:CAB11064.1; GSPDB:GN00066; SPDB:SPAC6B12.02c

C:Genetics:

A:Experimental source: strain 97zh-; cosmid c6B12

A:Gene: SPDB:SPAC6B12.02c

A:Map position: 1

C:Superfamily: Schizosaccharomyces hypothetical protein SPAC6B12.02c

Query Match 10.9%; Score 92.5; DB 2; Length 1888;

Best Local Similarity 23.1%; Pred. No. 81;

Matches 39; Conservative 25; Mismatches 76; Indels 29; Gaps 5;

Qy 1 TTVEFLINKDGEVSELKPHRVTVITIQNGKMSSTIVSE-EDFILPVYKGELEKGYQPD 59

Db 394 TSENFOLNVAANAVSTIPVTRTTTKKQKRFKVEKLPDLISEY----- 442

Qy 60 GWEISGFEGKQAGYVINSKDTFIRPVFKKIEKKKEENKPTPDVSKKQDPVNHSSOL 119

Db 443 -----GKAKPKLRFVARSSSHIF--KMIRRKQDSSKRYFSDKESPDQVIDYLS 492

Qy 120 NESHKEDLQREHSSQKSDS-TKDYATVLDKN-----NISKSTNN 161

Db 493 DWYSGKHLEVOOSHSHYKRPDSKSVGNIFVNSKSHNINAKTAMANN 541

## RESULT 31

A47705

triacylglycerol lipase (EC 3.1.1.3) - Staphylococcus epidermidis

C:Species: Staphylococcus epidermidis

C:Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999

C:Accession: A47705

R:Farrell, A.M.; Foster, T.J.; Holland, K.T.

J. Gen. Microbiol. 139, 267-277, 1993

A:Title: Molecular analysis and expression of the lipase of Staphylococcus epidermidis.

A:Reference number: A47705; MUID:93171870; PMID:8436947

A:Contents: 9

A:Accession: A47705

A&gt;Status: preliminary

A:Molecule type: DNA; protein

A:Residues: 1-688 &lt;FAR&gt;

A:Cross-references: GB:M95577; NID:g153021; PIDN:AAA19729.1; PID:g153022

A&gt;Note: sequence extracted from NCBI backbone (NCBIN:125632, NCBI:P:125633)

C:Superfamily: Staphylococcus triacylglycerol lipase

C:Keywords: carboxylic ester hydrolase

Query Match 10.8%; Score 92; DB 2; Length 688;

Best Local Similarity 25.0%; Pred. No. 28;

Matches 40; Conservative 28; Mismatches 66; Indels 26; Gaps 7;

Qy 9 NKDTGEVSELKPHRVTVITIQNGKMSSTIVSEEDFILPVYKGELEKGYQPDGMFI 68

Db 68 NKQVNEKSNVN-----SITENESLHNETPKNEDLI-----QOQKDSQDNKSSSVYEQ 115

Qy 69 KKDAG-VYINSKDTFIRPVFKIE-EKKEENKPTPDVSKKQDPV--NHSQNESHR 124

Db 116 NKENGAFQNHSEB---KQOEQVLEKHAENNOTLHASKAQSNEYKTKPSQLDNTAA 172

Qy 125 KEDLQREHSSQKSDSTKDVATVLDKNNISSKSTNNPNK 164

Db 173 KQEDSQKENSQDQTSKTDLL-----RATQNSQK 205

## RESULT 32

S05362

probable DNA-directed DNA polymerase (EC 2.7.7.7) - fungus (Ascobolus immerus) mitochondrion

C:Species: mitochondrion Ascobolus immerus

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000

C:Accession: S05362

R:Kempken, F.; Meinhardt, F.; Besser, K.

Mol. Gen. Genet. 218, 523-530, 1989

A:Title: In organello replication and viral affinity of linear, extrachromosomal DNA of

A:Reference number: S05362; MUID:90066356; PMID:2571821

A:Accession: S05362

A:Molecule type: DNA

A:Residues: 1-1202 &lt;KEM&gt;

A:Cross-references: EMBL:X15982; NID:g2933; PIDN:CAA34106.1; PID:g1370212

C:Genetics:

A:Genome: mitochondrion

A:Genetic code: SGC3

C:Superfamily: Ascobolus probable DNA-directed DNA polymerase  
C:Keywords: DNA binding; mitochondrion; nucleoside/transferase

Query Match 10.8%; Score 92; DB 1; Length 1202;

Best Local Similarity 22.5%; Pred. No. 53;  
Matches 42; Conservative 37; Mismatches 76; Indels 32; Gaps 9;

QY 6 FILNK-----DTGVSELSKPHRYVTIIONGKEMSTI---VSEEDF--ILPVYGELEK-- 54

DB 311 FVNVNAKIKKPTGNVRSIGGNVT-TLMDKTLIKTIAFLERDHTTWSYBGGDIDESK 369

QY 55 ----GYOPDGEWEISGEGKKAGYVINTSKPTFKPVFKIE-----EKKEEENK-PTF 103

DB 370 FPKGSLSPDKPKLKTIEGTVAAYTFPKKDIVVKKINKINFGDLPRXTMDLSKMPNL 429

QY 104 DVSKKDNPQVNSQUNESHRRKDLQREHSQKSDSTKVTATVLDKNNI-----SSK 156

DB 430 KLNKDKTSGEIRMTIKNNOSYDI--IGHMINDENVTTFNRAVANSIIRKIIFTVTDW 487

QY 157 STNNPN 163

DB 488 GNTNDPN 494

# RESULT 33

S38173 myosin-like protein MLP1 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YKR095W; protein YKR415

C:Species: Saccharomyces cerevisiae

C>Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 21-Jul-2000

R:Baladron, V.; Ballestra, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo, submitted to the Protein Sequence Database, March 1994

A:Accession: S38158

A:Reference number: S38158

A:Molecule type: DNA

A:Residues: 1-1875 <BAL>

A:Cross-references: EMBL:Z28320; NID:G486586; PID:G486587; MIPS:YKR095W

A:Experimental source: strain S288C

R:Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.

Mol. Gen. Genet. 237, 359-369, 1993

A:Title: A new yeast gene with a myosin-like heptad repeat structure.

A:Reference number: S31207; MUID:93247549; PMID:8483450

A:Accession: S31207

A:Molecule type: DNA

A:Residues: 1-300, 'A', 302-1875 <KOE>

A:Cross-references: EMBL:L01992; NID:G171958; PID:AAA34783.1; PID:G171959

C:Genetics:

A:Gene: SCD:MLP1

A:Cross-references: SGD:S0001803; MIPS:YKR095W

A:Map position: 11R

Query Match 10.8%; Score 92; DB 2; Length 1875;  
Best Local Similarity 25.0%; Pred. No. 87;  
Matches 48; Conservative 36; Mismatches 62; Indels 46; Gaps 10;

QY 7 ILNKDGEVSELKPHRYVTIIONGKEMSTI---ILPVYGE---LEKGYOF 58

DB 660 LNK---ELDLDVDSKSDISIKGKSSRIIAEERKLLSNLIDLTKAENDDLRRKFPD 716

QY 59 DGWEISGEGKDA-----GYVINSKDTFKPVFKIEK-----KEEENK 100

DB 717 ----LQNTILKODSKTHETLNEYVSCSKSIIVETELNLKEBOKRLVHLEKNLKOELNK 772

QY 101 PTFDVSKKKDNPOVNSQUNESHKEDL---QREHSQKSDSTKDVTA-----TVLDKN 151

DB 773 ----LSPEKDSRIWVTOQLQKEREDLLESTRSCCKKIDELDALSELKETSQKDH 828

QY 152 NISSKSTNNPN 163

DB 829 HIKOLEEDNNSN 840

# RESULT 34

D90093 heat shock protein 70KD [imported] - Guillardia theta nucleomorph

C:Species: nucleomorph Guillardia theta

A:Note: A nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont

C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 15-Jun-2001

C:Accession: D90093

R:Boydla, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Re

A:Title: The highly reduced genome of an enslaved algal nucleus.

A:Reference number: A99082; MUID:11323671; PMID:11323671

A:Accession: D90093

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-650 <DOU>

A:Cross-references: GB:AF165818; NID:G13794501; PID:AAK39876.1; GSPDB:GN00150

C:Genetics:

A:Gene: hsp70

A:Map position: 1

A:Genome: nucleomorph

C:Superfamily: heat shock protein 70

C:Keywords: nucleomorph

Query Match 10.8%; Score 91.5; DB 2; Length 650;

Best Local Similarity 23.6%; Pred. No. 29;

Matches 38; Conservative 19; Mismatches 51; Indels 53; Gaps 5;

QY 7 ILNKDGEVSELKPHRYVTIIONG---KEMSTI---ILPVYGELEKGYOPDGEWEI 63

DB 492 ILNVASCDSTGSKSKMTITNDKGRLSKEIRWVEAE---KYKNEDEKTRQ----- 541

QY 64 SGPEKKDAGYVINTSKDTFKPVFKIEKEEEN-----KPTDVSKKKDNPOVNSQ 118

DB 542 -----KIEAKNNLNYAVNINRTIRDEKDKIDENEXK 575

QY 119 LNES-----HKREDQREHSQKSDSTKDVATVLDK 150

DB 576 LBEKIRETLEVENNEDLEKEDYEKEKELKOWSNPIISK 616

RESULT 35

TJ2879

hypothetical protein C17F3.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T132879

R:Gatung, S.; Scheet, P.

submitted to the EMBL Data Library, January 1998

A:Description: The sequence of C. elegans cosmid C17F3.

A:Reference number: Z21240

A:Accession: T132879

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-253 <GAT>

A:Cross-references: EMBL:AF043692; PID:AA97531.1; GSPDB:GN00019; CESP:C17F3.3

A:Experimental source: strain Bristol N2; clone C17F3

C:Genetics:

A:Gene: CESP:C17F3.3

A:Map position: 1

A:Introns: 41/1

Query Match 10.7%; Score 91; DB 2; Length 253;

Best Local Similarity 32.6%; Pred. No. 11;

Matches 31; Conservative 7; Mismatches 33; Indels 24; Gaps 3;





A/Note: The genome of *Melanoplus sanguinipes entomopoxvirus*.  
 A/Reference number: Z20484; MWID:99102612; PMID:9847359  
 A/Accession: T28391  
 A/Status: preliminary; translated from GB/EMBL/DBD  
 A/Molecule type: DNA  
 A/Residues: 1-670 <AFO>  
 A/Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97713.1; PID:g4049753  
 C/Genetics:  
 A/Note: MSV230

Query Match 10.7%; Score 90.5; DB 2; Length 670;  
 Best Local Similarity 22.5%; Pred. No. 35;  
 Matches 36; Conservative 34; Mismatches 69; Indels 21; Gaps 5;

QY 25 VTIONKEMSSST-----VSEEDFILPYKGELEKGY----QPDGWEISGP 66  
 DB 33 VNEEEKQITITLKRNNEDKTEMCGVSEKFFQVLNNKASAEKYDSDSSIDSSONS 92  
 QY 67 EGRKAGYVNLISKDTFIRKPKIE-EKKEENKPTFDVSKKKNPQVNHSQL-NEHR 124  
 DB 93 DSDSDGVNIDSSQNSDKVNINKLENSQNSDSKVNIDSSQNSD-KVNINKLENSQ 151  
 QY 125 KEDLQREHSQKSDSTKDVATATVLDKNNISKSSTNNPNK 164  
 DB 152 SDSKNIDSSQNSDKVNIDSSQNSDKVNIDSSQNSDK 191

RESULT 39  
 S41552  
 probable transcription factor SPT7 - yeast (*Saccharomyces cerevisiae*)  
 N/Alternate names: protein YBR0739; protein YBR081C  
 C/Species: *Saccharomyces cerevisiae*  
 C/Date: 28-Jan-1994 #sequence\_revision 09-Sep-1994 #text\_change 20-Sep-1999  
 C/Accession: S41552; S45946; S45948; S40800; S45478; S5485; S57716  
 R/Gantheroff, L.; Dollard, C.; Tan, P.; Winston, F.  
 A/Title: The EMBL Data Library, July 1993  
 A/Reference number: S41552  
 A/Accession: S41552  
 A/Molecule type: DNA  
 A/Residues: 1-1332 <GAN>  
 A/Cross-references: EMBL:L22537; NID:g349189; PIDN:AAC37424.1; PID:g349190  
 R/Steenma, H.Y.; van der Aart, Q.J.M.  
 A/Title: The Protein Sequence Database, August 1994  
 A/Reference number: S45932  
 A/Accession: S45946  
 A/Molecule type: DNA  
 A/Residues: 1-1332 <STE>  
 A/Cross-references: EMBL:Z25950; NID:g536341; PIDN:CAA85026.1; PID:g536342; MIPS:YBR081C  
 R/Andre, B.; Czileplich, C.; Hein, C.; Jauniaux, J.C.; Urrestarazu, A.; Vissers, S.  
 A/Title: Submitted to the Protein Sequence Database, August 1994  
 A/Reference number: S45893  
 A/Accession: S45948  
 A/Molecule type: DNA  
 A/Residues: 1-835 <AND>  
 A/Cross-references: EMBL:Z25950; MIPS:YBR081C  
 R/Haynes, S.R.; Dollard, C.; Winston, F.; Beck, S.; Trowdale, J.; David, I.B.  
 A/Title: Nucleic Acids Res. 20, 2603, 1992  
 A/Reference number: S40800; MWID:92285152; PMID:1350857  
 A/Accession: S40800  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: DNA  
 A/Residues: 463-523 <HAY>  
 A/Cross-references: EMBL:M67651; NID:g172683; PIDN:AAA35087.1; PID:g172684  
 R/van der Aart, Q.J.M.; Barthe, C.; Dolignon, F.; Aigle, M.; Crouzet, M.; Steensma, H.Y.  
 A/Title: Yeast 10, 959-964, 1994  
 A/Reference number: S45462; MWID:95076715; PMID:7985423  
 A/Accession: S45478  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-624, 'LRGKKRI', 633-1332 <VAN>  
 A/Cross-references: EMBL:X76294

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993  
 R/Gantheroff, L.J.; Dollard, C.; Tan, P.; Winston, F.  
 A/Title: The *Saccharomyces cerevisiae* SPT7 gene encodes a very acidic protein important  
 A/Reference number: S54985; MWID:95229044; PMID:7713415  
 A/Accession: S54985  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: DNA  
 A/Residues: 1-1332 <GAN>  
 A/Cross-references: EMBL:L22537; NID:g349189; PIDN:AAC37424.1; PID:g349190  
 R/van der Aart, Q.J.M.  
 A/Title: Submitted to the EMBL Data Library, August 1995  
 A/Reference number: S59702  
 A/Accession: S59716  
 A/Molecule type: DNA  
 A/Residues: 1-1332 <VAM>  
 A/Cross-references: EMBL:X76294; NID:g974203; PIDN:CAA53940.1; PID:e264674; PID:g558360  
 A/Experimental source: strain S288C  
 C/Genetics:  
 A/Genes: SGD:SPT7  
 A/Cross-references: SGD:S0000285; MIPS:YBR081C  
 A/Map position: 2R  
 C/Superfamily: unassigned bromodomain proteins; bromodomain homology  
 C/Keywords: nucleus; transcription regulation  
 F/466-521/Domain: bromodomain homology <BRO>

Query Match 10.7%; Score 90.5; DB 2; Length 1332;  
 Best Local Similarity 24.5%; Pred. No. 76;  
 Matches 39; Conservative 26; Mismatches 57; Indels 37; Gaps 8;

QY 25 VTIONKEMSSSTVSEEDFILPYKGELEKGYQPDGWEISGEGKAGYVNLISKDTFI 84  
 DB 549 ITRIRADLEKSI---EDM-----EKDKYELDEEVEVAGSKRG---LNMGAHMLA 594  
 QY 85 K--PYFKTIEKKEENKPTD-----VSKKD-----NPQVHSQLNESH 123  
 DB 595 KENGKVEKSDSKTYDEAPTDKLTSTVPEGKEKDKTASTVTVAHNNKBEIKENG 654

QY 124 RKEDLQR--EHSQKSDSTKDVATATVLD-KNNISKSSTTN 160  
 DB 655 KNEBQWVESSKTEBSSKDAADAKDTEDGIDQDTAEN 693

RESULT 40  
 T28676  
 rhopty protein - *Plasmodium yoelii* (fragment)  
 C/Species: *Plasmodium yoelii*  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 01-Dec-2000  
 C/Accession: T28676; A45521  
 R/Sinha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.  
 A/Title: Mol. Biochem. Parasitol. 76, 329-332, 1996  
 A/Reference number: A45521; MWID:91101660; PMID:2270106  
 A/Accession: A45521  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 2260-2401 <KEE>  
 A/Cross-references: GB:M34281

Query Match 10.7%; Score 90.5; DB 2; Length 2401;  
 Best Local Similarity 27.4%; Pred. No. 1.5e+02;  
 Matches 43; Conservative 26; Mismatches 53; Indels 35; Gaps 6;

QY 26 VTIONKEMSSST-----VSEEDFILPYKGELEKGY---QPDGWEISGEGKAKDA 72

```
Db      924 TSKNHEKISDIRKNSLKIIOFSESESYINDI-KKELEKNTVLESQNNNTDINGVLSKIEN 982
Qy      73 GVINLSKDOTPIKPVFKKIEEKKEENKPTFDVSKKKONPOVNHSQLNESHKEDLQREE 132
Db      983 IY--NILKLNKIKKIIDKVKEYTDEIK-----NNKINAELSNSEKIITQLKENSLSKE 1035
Qy      133 HSQKSDSTKD-----VTATVLDKNNTSS 155
Db      1036 COSKIKSTIDNYSSECIKNITNLKTYIVNEKNNT 1072
```

Search completed: February 10, 2004, 10:58:35  
Job time : 12.8417 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:45 ; Search time 12.5638 Seconds  
(without alignments)  
1331.870 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897  
Sequence: 1 KIVVDFARNITTVKEFLINK.....ATVLDKNNISKSSTNNPNK 174

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	2140	2 F95074	serine proteinase,
2	894	99.7	2144	2 A97942	metalloproteinase
3	114.5	12.8	1038	2 JCS497	claustrin - chick
4	110	12.3	558	2 T18467	hypothetical prote
5	109	12.2	665	2 B71609	hypothetical prote
6	104	11.6	540	2 D86432	hypothetical prote
7	103.5	11.5	325	2 T18283	hypothetical prote
8	102.5	11.4	385	2 T20410	hypothetical prote
9	102	11.4	622	2 A90570	lipoprotein (impor
10	101.5	11.3	312	2 G81339	probable membrane
11	101.5	11.3	335	2 T37189	hypothetical prote
12	101	11.3	700	2 S67610	probable membrane
13	100	11.1	219	2 B72291	hypothetical prote
14	100	11.1	1875	2 S38173	myosin-like protei
15	99	11.0	211	2 T25911	hypothetical prote
16	99	11.0	1397	2 T10466	DNA topoisomerase
17	97.5	10.9	614	2 A84152	hypothetical prote
18	97.5	10.9	1345	2 S46817	hypothetical prote
19	97	10.8	644	2 T47835	hypothetical prote
20	97	10.8	988	2 T14188	hypothetical prote
21	96.5	10.8	456	2 T05612	hypothetical prote
22	96.5	10.8	867	2 T27136	hypothetical prote
23	96.5	10.8	871	2 T27135	hypothetical prote
24	96	10.7	491	2 C97267	hypothetical prote
25	96	10.7	1939	2 T18372	repeat organellar
26	96	10.7	2500	2 G71609	hypothetical prote
27	95.5	10.6	3724	2 T18427	hypothetical prote
28	95	10.6	348	2 T37271	cyclin II - human
29	95	10.6	1202	1 S05362	probable DNA-direc

30	94.5	10.5	277	2 D70214	surface lipoprotei
31	94.5	10.5	670	2 T28391	ORF MSV230 hypohe
32	94	10.5	210	2 T28771	hypothetical prote
33	94	10.5	553	2 T15094	hypothetical prote
34	94	10.5	947	2 T03795	ornithine decarbox
35	93.5	10.4	456	2 T03045	hypothetical prote
36	93.5	10.4	645	2 B89883	conserved hypothet
37	93.5	10.4	649	2 S42488	dnak-cype molecula
38	93	10.4	629	2 G96542	hypothetical prote
39	93	10.4	2464	1 QRMSP1	microtubule-associ
40	92.5	10.3	443	2 S66040	serine-type D-Ala-
41	92.5	10.3	571	2 B64469	hypothetical prote
42	92.5	10.3	1702	2 A41859	IGA-specific metal
43	92.5	10.3	1888	2 T39009	hypothetical prote
44	92.5	10.3	1997	2 F71607	hypothetical prote
45	92	10.3	688	2 A47705	DNA helicase II BR
					ctriacylglycerol 11

## ALIGNMENTS

```
RESULT 1
F95074
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)
C.Species: Streptococcus pneumoniae
C.Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C.Accession: F95074
R.Tetelish, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid,
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A.Authors: Loftus, B.O.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A>Title: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A.Reference number: A95000; MUID:21357209; PMID:11463916
A.Accession: F95074
A>Status: preliminary
A.Molecule type: DNA
A.Residues: 1-2140 <KIR>
A.Cross-references: GB:A905672; PIDN:NAK74791.1; PID:914972117; GSPDB:GN00164; TIGR:SP4
A.Experimental source: strain TIGR4
C.Genetics:
A.Gene: SP0641

Query Match      100.0%; Score 897; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 6.9e-57;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KIVVDFARNITTVKEFLINKDGEVSELPKPRVTYTIQNGKMSSTIVSEDFILPYK 60
DB      1933 KIVVDFARNITTVKEFLINKDGEVSELPKPRVTYTIQNGKMSSTIVSEDFILPYK 1992

QY      61 ELEKGYQDFGMEISGFEKSKDAGYVINSKDTFLKPVKKEEENKPTDVSXK 120
DB      1993 ELEKGYQDFGMEISGFEKSKDAGYVINSKDTFLKPVKKEEENKPTDVSXK 2052

QY      121 NPQVNSQLNSHREKEDIQREHSQKSDSYQVATATVLDKNNISKSSTNNPNK 174
DB      2053 NPQVNSQLNSHREKEDIQREHSQKSDSYQVATATVLDKNNISKSSTNNPNK 2106

RESULT 2
A97942
metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)
C.Species: Streptococcus pneumoniae
C.Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C.Accession: A97942
R.Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E,
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, F.; McAnnen, S.; M,
Y, P.; Sun, P.M.; Winkler, M.E.
J.Bacteriol. 183, 5709-5717, 2001
A.Authors: Yang, Y.; Young-Ballido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A>Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A.Reference number: A97872; MUID:21429245; PMID:11544234
```

A:Accession: A97942  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2144 <KUR>  
 A:Cross-references: GB:AE007317, PIDN:AAK9365.1, PID:g15458138, GSPDB:GN00174  
 C:Genetics:  
 A:Gene: prta  
 C:Keywords: hydrolase; serine proteinase

Query Match 99.7%; Score 894; DB 2; Length 2144;  
 Best Local Similarity 99.4%; Pred. No. 1.1e-56;  
 Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KIVVQKPARNTYKKEFLINKDGEVSELKPHRYVTITONGKEMSTIYSEEDFILPVYKG 60  
 DB 1937 KIVVQKPARNTYKKEFLINKDGEVSELKPHRYVTITONGKEMSTIYSEEDFILPVYKG 1996  
 QY 61 ELEKGYQFDGWEISGFEKGKADGYVINTSKDTFIKVPFKKEEKENKPTFDVSKKD 120  
 DB 1997 ELEKGYQFDGWEISGFEKGKADGYVINTSKDTFIKVPFKKEEKENKPTFDVSKKD 2056  
 QY 121 NPQVNSQINSHRKKEDLQREHSHQKSDSTKDVATVLDKNNISKSTNNPNK 174  
 DB 2057 NPQVNSQINSHRKKEDLQREHSHQKSDSTKDVATVLDKNNISKSTNNPNK 2110

## RESULT 3

JCS497  
 N:Alternates - chicken  
 C:Species: Gallus gallus (chicken)  
 C:Date: 07-Jul-1997 #sequence revision 12-Sep-1997 #text\_change 21-Jul-2000  
 C:Accession: JCS497, PC4334; S37561  
 R:Burg, M.A.; Cole, G.U.  
 J. Neurobiol. 25, 1-22, 1994  
 A:Title: Claustrian, an antiadhesive neural keratan sulfate proteoglycan, is structurally  
 A:Reference number: JCS497; PMID:94157526; PMID:7906711  
 A:Accession: JCS497  
 A:Molecule type: mRNA  
 A:Residues: 1-1038 <BUR1>  
 A:Cross-references: EMBL:X67778; NID:G406318; PIDN:CAA47988.1; PID:G406319  
 A:Accession: PC4334  
 A:Molecule type: protein  
 A:Residues: 79-83;299-432;485-502 <BUR2>  
 A:Experimental source: brain  
 C:Comment: This protein inhibits neural cell adhesion and neurite outgrowth in the nervo  
 C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; keratan sulfate  
 F:167-270/Region: cell attachment (R-G-D) motif  
 F:112,213,490/Binding site: carbohydrate (Aen) (covalent) #status predicted  
 F:152,249,440,793,820/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 12.8%; Score 114.5; DB 2; Length 1038;  
 Best Local Similarity 23.5%; Pred. No. 1.3;  
 Matches 47; Conservative 29; Mismatches 65; Indels 59; Gaps 6;

QY 2 IIVKQDARN---TTVEFLINKDGEVSELKPHRYVTITON----- 39  
 DB 514 VTQKDLTGNTASPAVVKQAKKORTDSKESLPAKATTTKDCCKRNLKKKMSLQSLVQO 573  
 QY 40 -----GKEMSTIYSEEDFILPVYKGELEKGYQFDGWEISGFEKG 79  
 DB 574 LEKPKLESKEKTPVKKKAVKRETKTIVAKCV-----TTKEQGLKSTSKQAS 625  
 QY 80 KDAQVINTSKDTFIKVPFKKEEKENKPTFDVSKKQNPQVNSQINSHRKKED 138  
 DB 626 EKQDVKPKTKKSKKVEVAKPEKKDEKPKKSKKPKPLI---KKEKKPKKEDI 682  
 QY 139 QRE-----EHSQKSDSTKDV 153  
 DB 683 KKEVKKEVKKEKKEKKEV 702

## RESULT 4

T18467  
 Hypothetical protein C0465c - malaria parasite (Plasmodium falciparum)  
 C:Species: Plasmodium falciparum  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2002  
 C:Accession: T18467  
 R:Lawson, D.; Bowman, S.; Barrell, B.  
 submitted to the EMBL Data Library, November 1998  
 A:Reference number: Z18937  
 A:Accession: T18467  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-558 <LAW>  
 A:Cross-references: EMBL:AL008970; NID:e1407852; PIDN:CAAL5610.1  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 84/1, 160/1  
 A>Note: C0465c

Query Match 12.3%; Score 110; DB 2; Length 558;  
 Best Local Similarity 29.3%; Pred. No. 1.3;  
 Matches 54; Conservative 23; Mismatches 47; Indels 60; Gaps 12;

QY 47 IYSEEDFILPVY-----KGELEKGYQFDGWEISGFEKGK---DAGYVINTSKDTFIKVP 97  
 DB 60 ILGFEDDILYEYCISQLKQSKKEK--KADGEEDKYNLAKKLINTLGFIGNKKSIDIFIEL 117  
 QY 98 FKKI---EKKKEE-----ENKPTFDVSK-KKQNPQVNSQIN-----SHRK 135  
 DB 118 LELLINERKKEHIDTLNENK--TNDIKYKNNENINENYNNENKDKSNKKEHVSION 176  
 QY 136 E-----DLQREH-----SQKSDSTK---DVTATVLDKNNISKSTTN 170  
 DB 177 EHNINNVNLMKKEKEYTDIQDRKFKRSLSQKSDSYKKRPFKRTSTIER-SLSNKRYSDE 235  
 QY 171 NPNK 174  
 DB 236 KTNK 239

## RESULT 5

B71609  
 Hypothetical protein PEB0680w - malaria parasite (Plasmodium falciparum)  
 C:Species: Plasmodium falciparum  
 C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
 C:Accession: B71609  
 R:Gardner, M.J.; Tettein, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
 ; Petrea, M.; Salzbeg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
 Science 282, 1126-1132, 1998  
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
 A:Reference number: A71600; PMID:99021743; PMID:9804551  
 A:Accession: B71609

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-665 <GAR>  
 A:Cross-references: GB:AE001410; GB:AE001362; NID:G3845245; PIDN:AACT1925.1; PID:G384524  
 A:Experimental source: clone 3D7  
 C:Genetics:  
 A:Gene: PEB0680w

Query Match 12.2%; Score 109; DB 2; Length 665;  
 Best Local Similarity 26.1%; Pred. No. 1.9;  
 Matches 49; Conservative 32; Mismatches 49; Indels 58; Gaps 11;

QY 5 KQPARNTYKKEFLINKDGEVSELKPHRYVTITONGKEMSTIYSEEDFILPVYKGELE 63  
 DB 114 KQNNNNNGTQKIEENKXINKSDL--HRQNELNLSGK-----NKNE 158  
 QY 64 KGYQFDGWEI--SGFEKGKADGYVINTSKDTFIKVPFKKEEKE-----ENKP 111  
 DB 159 KGGQ---DISNSNANKD-----VKSGVLELEKKEKESKSDHKVLENKK 202  
 QY 112 TFD---VSKKQNPQVNSQINSHRKKEDLQREHSHQKSDSTKDVATVLDKNNISK 166

```
Dh      57 HRTTISICK--RFSYVKIGDEEKLFRISKNGELLVNLNELEFDNPHIK--EGKILRKSMF 112
Qy      80 ---KDAGV-----INLSKOTFLKP-----FKK-----100
          ||:||| | | | :||
Dh     113 NHIKDSGYATNBEIEIPLFESCTLOCKETTAOTKRNSYKGNNINKLPREEEEEEEEEEE 172
Qy    101 --IEEKKEENKPTFDVSKKKNPOVNHSQLNESHKRE-----DLQRSEHSGKS DSTX 151
          ||:||| ||| ::::| | | | | | | | | | | |
Dh     173 EEEEBQEEVEKPTISEEEBETPAVSBEEKKEEEEEEEBETPAVSBEEKKEEBQEEDXEK 232
Qy    152 D 152
          233 D 233
Db
```

RESULT 8  
T20410  
hypothetical protein E02A10.2 - Caenorhabditis elegans  
C|Species: Caenorhabditis elegans  
C|Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
Accession: T20410  
R|Thomas, K.  
Submitted to the EMBL Data Library, October 1996  
Reference number: Z19271  
Accession: T20410  
Status: preliminary; translated from GB/EMBL/DDBJ  
Molecule type: DNA  
Residues: 1..385 <WILL>  
Cross-references: EMBL:X81053; PIDDN:CAB02877.1; GSPDB:GN00023; CESP:E02A10.2,  
Experimental source: clone E02A10  
Genetics:

[illegible]

```

Db      317  -KKEEKKKEEKKKE--EVEKKKE-----EKKKDEEPPKKEEKKKEEKKEDVEEKKSEKV 368
QY      158  LDK 160
          :|
Db      369  EEK 371

RESULT 9
A90570
lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CT1P)
C|Species: Mycoplasma pulmonis
C|Date: 24-May-2001 #sequence__revision 24-May-2001 #text_change 03-Aug-2001
C|Accession: A90570
R|Cambud, I.; Helling, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; M
Nucleic Acids Res. 29, 2145-2153, 2001
A|Title: The complete genome sequence of the murine respiratory pathogen Mycop
A|Reference number: A99512; M01D:21267165; PMID:1153084

```

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-622 <NUR>  
A;Cross-references: GB:AL445566; PID:q14089879; PIDN:CAC13638.1; GSPDB:GN00153  
A;Experimental source: strain UAB C11P  
C;Genetics:

A:Gene: MYPU 4650  
A:Genetic code: SGC3

Query Match 11.4%; Score 102; DB 2; Length 622;  
Best Local Similarity 24.9%; Pred. No. 5.7;  
Matches 48; Conservative 35; Mismatches 66; Indels 44; Gaps 11;

QY 9 RNTTVEFIL-NQDTGEVSEL-----KPHRV---TVTIQNG--KEM--STI--V 48  
DB 26 KDSAKNPFVYKSSDLISISKEFPKRNNDKOKIIEGVITINSETKEIYKKEITLNL 65  
QY 49 SEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGVYINLSKDTFIKVPFKKIEEKEE 108  
DB 86 SEDNIIIFSLININNAKFFQDDEF-----VSKDEKFKIKFOEINFSQTEQ 129  
QY 109 NKRTFVSKKKD---NPQVNSQLNES--HKEDIQREHSQKSDTQVATVLDKX- 161  
DB 130 -KITDNISSEKEDKNKPKDNENSNNSSDQKNDELQKNNSDKLNDVDEKANKENSNS 188  
QY 162 NISSKSTNNPNK 174  
DB 189 NDSKEKNDERTNK 201

## RESULT 10

G81339  
probable membrane protein Cj0692c [imported] - Campylobacter jejuni (strain NCTC 11168)

C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C:Accession: G81339

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrel

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: A81250; PMID:20150912; PMID:10688204

A:Accession: G81339

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-312 <PAR>

A:Cross-references: GB:AL139076; GB:AL111168; NID:G6968128; PIDN:CAM72966.1; PID:G696814

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj0692c

Query Match 11.3%; Score 101.5; DB 2; Length 312;  
Best Local Similarity 25.0%; Pred. No. 2.8;  
Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;

QY 15 EFLINDTGEVSELKPHRVTVTIQNGKEMSSSTI---VSEEDFILPVYK-----GELE 63  
DB 5 DFIKERQNTIRKMLKFSR---AINQGRPLDDLDRLDISDDILRRFKKKTTPKFLIED 61  
QY 64 KGYQFDGWEISGFEKKDAGVYINL---SKDTFIKVPFKKIEEKEEENKPT---FDVSK 117  
DB 62 EEEYSGHTKTSKNLYLKED---LINVKLEEKOSLAKIFSKRRERRENNKTKKNLFSR 118  
QY 118 KKNP-----QVNSQLNESHRKEDLQREHSQKSDSTKV--TATVLDKNNISK--STT 169  
DB 119 KKNKNIKTQTKQIOTKSNQATTTQTKQEKELTNSIEKIQTETKIQKPLEIKLVDK 178  
QY 170 NNPN 173  
DB 179 NQPN 182

## RESULT 11

T37189

hypothetical protein CO2H7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 18-Feb-2000

C:Accession: T37189

R:Leimback, D.; Minx, M.

submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid CO2H7.

A:Reference number: Z20523

A:Accession: T37189

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-535 <LEI>

A:Cross-references: EMBL:U49945; PIDN:AAC47924.1; GSPDB:GN00029; CESP:CO2H7.1

A:Experimental source: strain Bristol N2; clone CO2H7

C:Genetics:

A:Gene: CESP:CO2H7.1

A:Map position: X

A:introns: 47/3; 149/3; 304/2; 347/3; 458/3

Query Match 11.3%; Score 101.5; DB 2; Length 535;  
Best Local Similarity 21.0%; Pred. No. 5.2;  
Matches 39; Conservative 31; Mismatches 71; Indels 45; Gaps 6;

QY 4 VKDFANTTVKEFIINK-----DTGEVSELKPHRVTVTIQNGKEMSSSTIIVSEEDFILPVY 58  
DB 55 IKSAGNDKNTKTAFLDKLIKIIDGSLKNVKAII---SGKDAEET----- 98  
QY 59 KGELEKGYQFDGWEISGFEKKDAGVYINLSKDTFIKVPFKKIEEKEEENKPTFDVSK 118  
DB 99 ---NKMQLMTNMTSPNSRGTG-----EKKKKKKVKKEDKKGDEEKST---TKK 144  
QY 119 KNPQVNSQLNESHRKEDLQREHSQKSDSTK-----DYATVLDKNNISKST 168  
DB 145 RSKKETHEKRSKSKSAEKEKKKSSSKENHNSDRSEKSKSEKSKKKEKST 204  
QY 169 TNNPNK 174  
DB 205 TDEKPK 210

## RESULT 12

S67610  
probable membrane protein YDL074c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein D2483

C:Species: Saccharomyces cerevisiae

C:Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 19-Apr-2002

C:Accession: S67610

R:Wambut, R.; Wedler, H.; Wedler, E.; Scharfe, M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67608

A:Accession: S67610

A:Molecule type: DNA

A:Residues: 1-700 <WAM>

A:Cross-references: EMBL:Z74122; NID:G1431087; PID:E253213; PID:G1431088; GSPDB:GN00004;

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:BRE1; MIPS:YDL074C

A:Cross-references: SGD:S0002232

A:Map position: 4L

C:Keywords: transmembrane protein

F:69-85/Domain: transmembrane #status predicted <TM>

Query Match 11.3%; Score 101; DB 2; Length 700;  
Best Local Similarity 22.4%; Pred. No. 7.7;  
Matches 46; Conservative 39; Mismatches 60; Indels 60; Gaps 10;

QY 5 KDFANTTVKEFIINK-----DTGEVSELKPHRVTVTIQ-----NGKEMSSST 46  
DB 379 KDVVIRIARTADLLKINILAEKSKTEVLSDIQ-HAIDLKEQWTKIDQSRNDTKSSST 437  
QY 47 IVSEEDFILPVYKGELEKGYQFDGWEISGFEKK-----DAGTV 85  
DB 438 ---ODALIKETIQ-DLEKGF---ELSDLTFRKYSIINHSVISKLVEKTKADQKXF 488  
QY 86 INL-SKDTF---IKVFKKIEEKEEENKPTFDVSKKKNPNQVNSQLNESHRKEDLORE 141  
DB 489 AAMRSKDSITLIEKTLSSLSKSNEL-----ILQKSDRLLOOKIGNKHQDLDSQ 541  
QY 142 EHSQKSDSTQVATVLDKNNISK 166





A:Residues: 1-1397 <CHE>  
 A:Cross-references: EMBL:X79345; NID:g9934807  
 C:Genetics:  
 A:Gene: TOP011  
 A:Map position: 14  
 C:Superfamily: eukaryotic type II DNA topoisomerase; phage T4 DNA topoisomerase (ATP-hyd  
 C:Keywords: ATP; DNA binding; isomerase; nucleus

Query Match 11.0%; Score 99; DB 2; Length 1397;  
 Best Local Similarity 22.2%; Pred. No. 24;  
 Matches 44; Conservative 45; Mismatches 67; Indels 42; Gaps 9;

QY 1 KIVKQFAR-----NTTKEFILNKDTGEVSEIKPHRYVTITONGKMSSTIVSE 51  
 DB 1092 KVLVEELKRGYQPYKDNKIKKEEIFEGELDAAD-NPE-----DNEELIGITVVKY 1144  
 QY 52 DFLI--PVKGELEKGYQPDGMEISGFEKKDAGYINISKOTFIKPVFKIIE----- 103  
 DB 1145 DYLLSMPIFSLTLEK--VEDLLTQLEKERELEIIRNITVEFMWLKDIEKVEALEFQR 1201  
 QY 104 -----KKEENKPTPDVSKKKNQPVNHSQLNESHKEDLQREHSQ---KSDTKDVA 155  
 DB 1202 NVELSNRESNK--FKYARKQ-----GPSMVKKKKKKKLSDESBGSDTSSEFLVN 1254  
 QY 156 TVLDKNNISKSSTNNPN 173  
 DB 1255 TLNKKNTNKKTTSSNN 1272

RESULT 17  
 A84152  
 hypothetical protein BH4017 [imported] - Bacillus halodurans (strain C-125)  
 C:Species: Bacillus halodurans  
 C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001  
 C:Accession: A84152  
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and  
 A:Reference number: A83650; PMID:20512582; PMID:11058132  
 A:Accession: A84152  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-614 <STO>  
 A:Cross-references: GB:AP001520; GB:BA000004; NID:g10176401; PIDN:BAB07736.1; GSPDB:GN00  
 A:Experimental source: strain C-125  
 C:Genetics:  
 A:Gene: BH4017

Query Match 10.9%; Score 97.5; DB 2; Length 614;  
 Best Local Similarity 27.3%; Pred. No. 12;  
 Matches 41; Conservative 26; Mismatches 44; Indels 39; Gaps 8;

QY 26 SELKPHRYVTITONGKMSSTIVSE-----EDFILPVKGELEKGYQPDGM---EISG 75  
 DB 386 SEBKPPE--TYTLQTALQM--TPIVNEVSPQTRSEFL-----ARKAQLDGMADVSKVY 435  
 QY 76 FEKKDAGYINISKOTFIKPVFKIIEKK--EEENKPTPDVSKK---DNQVNHSQL 129  
 DB 436 FAKGNIASQL-----GKIEEKELQDKYNNYITDPFKKEVNVQSPKISTAL 482  
 QY 130 NESHKEDLQREHSQKSDTKDVTATLD 159  
 DB 483 GKVIQIVDLGRKRYHKGKEDLERLSKSQIE 512

RESULT 18  
 S46817  
 hypothetical protein YHR080C - Yeast (Saccharomyces cerevisiae)  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 19-Apr-2002  
 C:Accession: S46817  
 R:Favell, T.  
 submitted to the EMBL Data Library, June 1994

A:Description: The sequence of S. cerevisiae cosmid 9205.  
 A:Reference number: S46795  
 A:Accession: S46817  
 A:Molecule type: DNA  
 A:Residues: 1-1345 <FAV>  
 A:Cross-references: EMBL:U10556; NID:g500825; PIDN:AA68895.1; PID:g500838; MIPS:YHR080C  
 C:Genetics:  
 A:Cross-references: SGD:S0001122  
 A:Map position: 8R  
 C:Keywords: transmembrane protein

Query Match 10.9%; Score 97.5; DB 2; Length 1345;  
 Best Local Similarity 27.0%; Pred. No. 29;  
 Matches 33; Conservative 18; Mismatches 38; Indels 33; Gaps 6;

QY 59 KGELEKGYQPDGMEISGFEKK-DAGYVYNISKOTFIKPVFKIIEKKEENKPTPDVSK 117  
 DB 1109 KCAIEKG-----SVEGQVSVDMYLSERDIT-----SRKSKKPVYKRYWK 1149  
 QY 118 KKDNPVNHSQLNESHKEDLQREHSQKSDTKDVTATLD--KNNISKSSTN---NP 172  
 DB 1150 SHDKRPFPSKY-----QKSESRKSDNKOILTHILDVQNNPSSSEIFNMKLLSP 1201  
 QY 173 NK 174  
 DB 1202 QK 1203

RESULT 19  
 T47835  
 hypothetical protein T209.90 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
 C:Accession: T47835  
 R:Yakutana, G.; Fatmann, B.; Dauner, D.; Sterr, W.; Holland, R.; Weichselgartner, M.;  
 submitted to the Protein Sequence Database, February 2000  
 A:Reference number: Z24475  
 A:Accession: T47835  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-644 <NYA>  
 A:Cross-references: EMBL:AL138658  
 A:Experimental source: cultivar Columbia; BAC clone T209  
 C:Genetics:  
 A:Map position: 3  
 A:Introns: 158/2; 329/3  
 A:Note: T209.90

Query Match 10.8%; Score 97; DB 2; Length 644;  
 Best Local Similarity 22.9%; Pred. No. 14;  
 Matches 43; Conservative 29; Mismatches 72; Indels 44; Gaps 7;

QY 8 ARNTTYKEFILNKD--TGEV--SELKPHRYVTI-----IQNGKMSSTIVSE--- 51  
 DB 476 SRTSKELAVNVAKDTKTKGAKNNIKQDPTKTESSDDDDDEKENSKTIEKTVADKKSV 535  
 QY 52 -PFIIPVYVYGELEKGYQPDGMEISGFEKKDAGYINISKOTFIKPVFKIIEKKEENK 110  
 DB 536 ADPLKRIKNSPOKKG-----ETSKNQKNDGNV-----KENDHQ 572  
 QY 111 PTFDVSKKDNPVNHSQLNESHKEDLQREHSQKSDS---TKDVTATVLDKNNISK 166  
 DB 573 KSDGNVKKENSKVPRRLRSSTGKKVVEVNNSSKSRKQTKETAEVATGKGRREG 632  
 QY 167 STTNPNK 174  
 DB 633 KDDKQPRK 640

RESULT 20  
 T1188  
 hypothetical protein T28D5.30 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)

RESULT 22  
T27136  
hypotheoretical protein Y53C12B.3a - *Caenorhabditis elegans*  
C:\Spectra: *Caenorhabditis elegans*

RESULT 24  
C97267  
hypothetical protein CAC2985 [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C:Accession: C97267  
R:Rollins, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C1C  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: C97267  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-491 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK80926.1; PID:G15026040; GSPDB:GN00168  
A:Experimental source: Clostridium acetobutylicum ATCC824  
C:Genetics:  
A:Gene: CAC2985

Query Match 10.7%; Score 96; DB 2; Length 491;  
Best Local Similarity 24.6%; Pred. No. 12;  
Matches 51; Conservative 33; Mismatches 69; Indels 54; Gaps 11;

QY 1 KIVKDFPANTVKEFTLNKDTGEVSELKPHRYVTITQNGKMSSTIVSEEDFILPVYKG 60  
DB 77 KINIDYGRSEISREF-----EESDVAGSGIADRVVGASISOPF---EDNIYVWSG 126  
QY 61 ELEKGVQPDGWEISG-----FEGKKDAGVYINLSKOTFIKPV---FKIE-----E 103  
DB 127 -----FNASDVSQSKYSPVVEKKEAPV---KREVEKPKKODKKEVKEKSTTD 175  
QY 104 KKEENKPTFD-----VSKKDN-PQVNSQLE-----SHRKEDLOREHSQKSD----- 148  
DB 176 KKENLNRIIPDKYEEKIEKKNHEIEKKEVNTKKEDPLKKEETNVDDKKEPR 235  
QY 149 -STKDVATVLDKNNISSKSTTNPNK 174  
DB 236 VEKKEETEVEKKEIEVKKEETKESR 262

## RESULT 25

T18372  
repeat organellar protein - Plasmodium chabaudi  
C:Species: Plasmodium chabaudi  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18372  
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.  
Mol. Biochem. Parasitol. 94, 185-196, 1998  
A:Title: A Plasmodium chabaudi protein contains a repetitive region with a predicted spe  
A:Reference number: Z18922; MUID:98418765; PMID:9747969  
A:Accession: T18372  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1939 <WER>  
A:Cross-references: EMBL:U43145; NID:G1151157; PID:G1151158; PIDN:AA63403.1

Query Match 10.7%; Score 96; DB 2; Length 1939;  
Best Local Similarity 22.6%; Pred. No. 57;  
Matches 47; Conservative 46; Mismatches 73; Indels 42; Gaps 10;

QY 1 KIVKDFPANTVKEFTLNKDTGEVSELKPHRYVTITQNGKMSSTIVSEEDFIL 55  
DB 339 KLEIKDLSQSLVEKEREIFEIKNEYDCKINMKKLSI---NDKQIDNTVLSSEKIN 395  
QY 56 PLYK-----GELEKYOQPDGWEISGFEKKDAGY-----VINLSKOTFIKPVK 99  
DB 396 KILKEKETELNEITHKKYKLEIETIKNELNEKEKELEKKNKAHTVEVTNLTKE--IKLLEK 453  
QY 100 KIEEKE-----EENKPTFDVSKKQNPQVNSQLE-----HRKEDLOREHSQKSD 148  
DB 454 KTEDEKGEKNELENNQSLKLNKEDKNIKNENTELNDKISSLSNEVVIANKDKOTLGN 513  
QY 149 STKDVATVLDKNNISSKSTTNPNK 174  
DB 514 DIKTLDLI---NNLKNINTSDNKNK 538

## RESULT 26

G71609  
hypothetical protein PFB0650w - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
C:Accession: G71609  
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
Pertes, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
Science 282, 1126-1132, 1998  
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
A:Reference number: A71600; MUID:99021743; PMID:9804551  
A:Accession: G71609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2500 <GAR>  
A:Cross-references: GB:AE001408; GB:AE001362; NID:G3845238; PIDN:AACT1919.1; PID:G384524  
A:Experimental source: clone 3D7  
C:Genetics:  
A:Gene: PFB0650w

Query Match 10.7%; Score 96; DB 2; Length 2500;  
Best Local Similarity 26.1%; Pred. No. 77;  
Matches 41; Conservative 30; Mismatches 54; Indels 32; Gaps 7;

QY 29 KHRVTVITQNGKMSSTIVSEEDFILPVYKGELEKGVQPDGWEISGFEKKDAGVINTL 88  
DB 2173 KPYKIT---ENKK-----NEGNEILKYSLENEKNKYDKEQENCLDQTCQNVVT 2223  
QY 89 -----SKDTFIKVFKEIEKEEENKPTFDVSKKQNPQVNSQLE-----SHRKED 137  
DB 2224 KKKNNLDNKKSPSNIVKYKLEEEKSDKRD---DKQNDTEKNNLDNKKSPSNIVKYK 2280  
QY 138 LOREHSQKSDTKDVTATVLDKNNISSKSTTNPNK 174  
DB 2281 LEEER---KSDDKRD-----DKQNDTEKNNLDNKK 2308

## RESULT 27

T18427  
hypothetical protein C0335C - malaria parasite (Plasmodium falciparum)  
C:Species: Plasmodium falciparum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000  
C:Accession: T18427  
R:Lawson, D.; Bowman, S.; Barrell, B.  
submitted to the EMBL Data Library, August 1997  
A:Reference number: Z18935  
A:Accession: T18427  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3724 <LAW>  
A:Cross-references: EMBL:Z98547; NID:E1325376; PID:E1325379; PIDN:CA11104.1  
C:Genetics:  
A:Introns: 307/1; 1545/2  
A>Note: C0335C

Query Match 10.6%; Score 95.5; DB 2; Length 3724;  
Best Local Similarity 19.1%; Pred. No. 1.3e+02;  
Matches 40; Conservative 36; Mismatches 66; Indels 67; Gaps 7;

QY 1 KIVKDFPANTVKEFTLNKDTGEVSELK----- 29  
DB 982 KILKDMKKKQEQEFKTDETGSIQSHKIKKYNNGEEKHDKNNEEKNILYDENQYVS 1041  
QY 30 ---PHRYVTITQNGKMSSTIVSEEDFILPVYKGELEKGVQPDGWEISGFEKKDAGYVI 86  
DB 1042 LYSCHKIEDQIDHSIQNICDENN-IEQINEEKSKGVRISGIDM---ENKND----- 1092  
QY 87 NLSKOTFIKPVFKKIEKKEEENKPTFDVSKKQNPQVNSQLESHRKEDLOREHSQ 145  
DB 1093 -----MENKNDMEKN--DMEKKNDIEKQNDMEKKNDEKKNDEKKNDE 1136

Oy 146 KSDSTKDVATVLDKNISKSSTNNPK 174  
Db 1137 KKNDME-----NENNENKSDIENENK 1158

## RESULT 28

137271  
Cyllicin II - human  
C:Species: Homo sapiens (man)  
C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 21-Jul-2000  
C:Accession: 137271; S52774  
R:Hees, H.; Heid, H.; Zimbelmann, R.; Franke, W.W.  
Exp. Cell Res. 218, 174-182, 1995  
A:Title: The protein complexity of the cytoskeleton of bovine and human sperm heads: the  
A:Reference number: 137271; MUID:95255491; PMID:7737358  
A:Accession: 137271  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-348 <HE>  
A:Cross-references: EMBL:Z46788; NID:g758586; PIDN:CAA6752.1; PID:g758587

Query Match 10.6%; Score 95; DB 2; Length 348;  
Best Local Similarity 30.9%; Pred. No. 9.4;  
Matches 38; Conservative 20; Mismatches 39; Indels 26; Gaps 7;

Oy 59 KGELEKGYQFQGWELSGFEGKKDAGYVNLKDTIFKVPFKIEKKKEENKPTF--DV 115  
Db 205 ESEGGK-----GTEKDSKKKKDS---KKGDSAKIELQAVKADKDEGDKKANKGDE 256

Oy 116 SK--KKDPOVNSOLN-----ESHKEDLQREHSQKSDTKD---VTATVLDKNII 163  
Db 257 SKDAKKDAKEIKKKKKKSDSTSDSKDVKKE---SKDADKDAKKVAKKTEKESA 313

Oy 164 SSK 166  
Db 314 DSK 316

## RESULT 29

S05362  
probable DNA-directed DNA polymerase (EC 2.7.7.7) - fungus (Ascombolus immerus) mitochon  
C:Species: mitochondrion Ascombolus immerus  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 16-Jun-2000  
C:Accession: S05362  
R:Kempken, F.; Meinhardt, F.; Esser, K.  
Mol. Gen. Genet. 218, 523-530, 1989  
A:Title: In organello replication and viral affinity of linear, extrachromosomal DNA of  
A:Reference number: S05362; MUID:90066356; PMID:2573821  
A:Accession: S05362  
A:Molecule type: DNA  
A:Residues: 1-1202 <KEM>  
A:Cross-references: EMBL:X15982; NID:g2933; PIDN:CAA34106.1; PID:g1370212  
C:Genetics:

A:Genome: mitochondrion  
A:Genetic code: SGC3  
C:Superfamily: Ascolobus probable DNA-directed DNA polymerase  
C:Keywords: DNA binding; mitochondrion; nucleotidyltransferase

Query Match 10.6%; Score 95; DB 1; Length 1202;  
Best Local Similarity 22.8%; Pred. No. 39; Indels 32; Gaps 9;  
Matches 44; Conservative 37; Mismatches 80;

Oy 10 NTTVEKFLINK-----DTGEVSELKPRVTVTIQNGKMSSTI---VSEEDF--ILPVYKG 60  
Db 305 NTFQFVFNNAKIKPTGNVRSIGQNT-TLTDKETILKLAIFLEREDITWVSDEG 363

Oy 61 ELK-----GYQFGWELSGFEGKKDAGYVNLKDTIFKVPFKIE-----EKSEB 108  
Db 364 DIDESKPRKGLSPDFKPLKTEGKRYANYTPPIKDIIVVKNINKNFNGSLDEKPTDL 423

Oy 109 NK-PFFDVSKKDNQVNSQLNESHKEDLQREHSQKSDTKDVTATVLDKNII----- 163  
Db 424 SKPWKLKDKDKTSGEIKRTIKKKNOSYDI--IGHMIINDGENVITFRRAVDNSIIKIF 481

Oy 164 ---SSKSTNNPN 173  
Db 482 TVTDSMGNTNDPN 494

## RESULT 30

D70214  
Surfactant lipoprotein P27 homolog - Lyme disease spirochete plasmid A/1p54  
C:Species: Borrelia burgdorferi (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C:Accession: D70214  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Jachigra, R.; White  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.  
A:Reference number: A70100; MUID:98065943; PMID:9403685  
A:Accession: D70214  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-277 <KLE>  
A:Cross-references: GB:A0000790; NID:g269024; PIDN:AAC6239.1; PID:g2690238; TIGR:BBA60  
A:Experimental source: strain B31  
C:Genetics:

A:Genome: plasmid.  
Query Match 10.5%; Score 94.5; DB 2; Length 277;  
Best Local Similarity 20.9%; Pred. No. 7.8;  
Matches 34; Conservative 28; Mismatches 62; Indels 39; Gaps 4;

Oy 4 VKDFARNTTVKRFINKDTGSEVSELKPRVTVTIQNGKMSSTIVSEDFILPVYKGELE 63  
Db 142 LKNEKEIKNIENFPQNDLFLVTLKDKNNNTINIMLPNDIQKPDYILKDKDKTK 201

Oy 64 KGYQFGWELSGFEGKKDAGYVNLKDTIFKVPFKIEKKKEENKPTFVSKKKDPQ 123  
Db 202 KG-----TGEK-----YINPIYR-----FQIKKKDYHS 225

Oy 124 VNSQLNESHKEDLQREHSQKSDTKDVA---TVLDKNII 163  
Db 226 IDYKVTISSEKTELDLPHQVQMNKPFKIDITIDNLN 268

## RESULT 31

T28391  
ORF MSV230 hypothetical protein - Melanoplus sanguinipes entomopoxvirus  
C:Species: Melanoplus sanguinipes entomopoxvirus  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
C:Accession: T28391  
R:Afonso, C.L.; Tuijman, E.R.; Lu, Z.; Oma, E.; Kutch, G.F.; Rock, D.L.  
J. Virol. 73, 533-552, 1999  
A:Title: The genome of Melanoplus sanguinipes entomopoxvirus.  
A:Reference number: Z20484; MUID:99102612; PMID:9847359  
A:Accession: T28391  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-670 <AFQ>  
A:Cross-references: EMBL:AF063866; NID:g4049647; PIDN:AAC97713.1; PID:g4049753  
C:Genetics:

A:Note: MSV230  
Query Match 10.5%; Score 94.5; DB 2; Length 670;  
Best Local Similarity 21.9%; Pred. No. 22;  
Matches 42; Conservative 39; Mismatches 82; Indels 29; Gaps 6;

Oy 3 VKDFARNTTVKRFINKDTGSEVSELKPRVTVTIQNGKMSSTI-----VS 49  
Db 9 ILDKVINMKTEKF-----KLKRYDNVNVFEKKQIISTLTKENNFDKTEKCVS 60

Oy 50 EEDFILPVYKGELEKGY-----QFDGWEISGFEGKKDAGYVNLKDTIFKVPFKIE-E 103  
Db 1137 KKNDME-----NENNENKSDIENENK 1158



[illegible]

RESULT 36  
E898883  
conserved hypothetical protein SA0576 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: E898883  
R:Kuroda, M., Ohta, T., Uchiyama, I., Babe, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguma, A., Mizutani-Ui, Y., Kobayashi, N., Sawano, T., Inoue, R., Kaito, C., Sekimizu, K., Shiba, T., Hattori, M., Ogasawara, N., Hayashi, I., Hiratsuku, K.  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: E898883  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-645 <KUR>  
A:Cross-references: GB:BA000018; PID:gl3700929; PIDB:BA042225.1; GSFDB:GN00149  
A:Experimental source: strain N315  
C:Gene: SA0576

	Query Match	10.4%	Score 93.5;	DB 2;	Length 645;
	Best Local Similarity	21.9%	Pred. No. 24;		
	Matches	49;	Conservative	34;	Mismatches 74; Indels 67; Gaps 11;
OY	11	TTVKEFLINKDTGE-VSELKPHRYTV--TTONGEMSTSIYSEEDFIIPYKGBLE-KGY	66		
Db	327	SAITEFQVNCPTNEMETDLQDTKYVVYESVENNESMDTFVKH-----PIKTGLMNGKKY	381		
OY	67	QP-----DGMESGFEGKK-----	81		
Db	382	MMETTNDIYNKDPMVEGOVRATTSKDAKNNTRTIIIFYEGKITLYAIVKHAKTIIDYD	441		
OY	82	AGYVINL-SKDTFIKPVFKJIEKKKEENKEPTDV-----SKKDNPQVNSQLINESHRK	135		
Db	442	GQYHVRIADVKAFTANTDKNKKEQQONSAAKEATPATPSPKPFPSPVESKQKODSOKD	501		
OY	136	EDLQ----REHSQKSDSTKVVT-ATVADKNNISKSQTNNPNK	174		
Db	502	DNKOLPSVEKENADASSEGCKTPATPKPTTKGEVSSSTLT--PTK	543		

RESULT 37  
542488  
dnaK-type molecular chaperone hsp70 - *Pyrenomonas salina* nucleomorph  
N.Alternate names: heat shock protein 70  
C.Species: nucleomorph *Pyrenomonas salina*  
A.Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont  
C.Date: 13-Jan-1995 #sequence \_revision 27-Jan-1995 #text\_change 20-Sep-1999  
C.Accession: 542488; S45576  
R.Resting, S.; Hofmann, C.U.B.  
Submitted to the EMBL Data Library, March 1993

```

A:Description: Smallest known eukaryotic genomes encode a protein gene: towards underst
A:Reference number: S42488
A:Accession: S42488
A:Molecule type: DNA
A:Residues: 1-649 <RES>
A:Cross-references: EMBL:X72621; NID:9461335; PIDN:CAA51197.1; PID:9461336
R: Hofmann, C.U.B.; Rensing, S.A.; Haebler, M.M.; Martin, W.F.; Mueller, S.B.; Couch, J.
Mol. Genet. 243, 600-604, 1994
A:Title: The smallest known eukaryotic genomes encode a protein gene: towards an underst
A:Reference number: S45576; MUID:94268506; PMID:8208251
A:Accession: S45576
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-24, 'O', 26-91, 'H', 93-578 <HO>
A:Cross-references: EMBL:X72621; NID:9461335
A:Note: The sequence is revised in Genbank entry PSHS70, release 111.0, (PIDN:CAA51197.
C:Genetics:
A:Gene: hsp70
A:Genome: nucleomorph
C:Function:
A:Description: involved in protein folding and assembling/disassembling of protein comp
C:Superfamily: heat shock protein 70
C:Keywords: ATP; molecular chaperone; nucleomorph
Query Match 10.4%; Score 93.5; DB 2; Length 649;
Best Local Similarity 24.5%; Pred. No. 25;
Matches 38; Conservative 26; Mismatches 50; Indels 41; Gaps 7;
17 ILNKDTGEVSEILKPHRVTVTIQNG--KEMSSTVISEDFILPYKCELEKGYQFDGWEI 73
||| : : : : : ||| : : : : :
DB 491 ILNVSASDKSTGKNKIKITITNDKGRLSKETIERAVEEAF---KYKTEDEK----- 537
74 SGFEKGDAD-----GYVINS--KQFFIKVPFKIIEKEEENKKPFVDVSKKDNQVN 125
: : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 538 --LDRKLEAKNSLDENVYNIIRNTVRD-----EKLEKIQEEDKSIIEKVE----- 582
126 HSQLNESHKRDLOREHSQKSDSTKDVATVLDK 160
||| : : : : : : : : : : : : : : : : : : : : :
DB 583 --VLEFTEITNEDLEKEVEKEKEKELKNFANPIIK 615

```

Query Match	10.4%	Score 93.5	DB 2	Length 649
Best Local Similarity	24.5%	Pred. 100		
Matches	38	Conservative	26	Mismatches 50
				Indels 41
				Gaps 7
QY	17	ILNKGTVSEILKPRVTVTIQNG--KEMSSITVSEEDFILPYKGELEKGYQFDGWEI	73	
DB	491	ILNVSASDKSTGKSNKITITNDKGRLSKEETIERVVEAE--KYKTEDEK-----	537	
QY	74	SGFEQKDA-----GYVINS--KOTFIKVPFKIIEKEEENKPFVDVSKKDNQVN	125	
DB	538	--LDRKLEAKNSLBNVYNIIRNTVRD-----EKLEKIQEEDKSIIEKVE-----	582	
QY	126	HSQLNESHKEDLQREHSQKSDSTKDVATVLDK	160	
DB	583	--VLEFIETNEDLEKEVEKEKELKNFANPIISK	615	

RESULT 38  
 G96542  
 hypothetical protein F17J6.14 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: G96542  
 R:Thelloglis, A.; Ecker, J.R.; Palm, C.J.; Federpohl, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Hultzer, L.  
 Nature 408, 815-820, 2000  
 C:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C  
 A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzalli,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: G96542  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-629 <STD>  
 A:Cross-references: GB:A005173; NID:G1054631; PIDN:AAG27876.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F17J6.14  
 A:Map position: 1

```
Query Match      10.4%; Score 93; DB 2; Length 629;  
Best Local Similarity 19.6%; Pred. No. 26;  
Matches    40; Conservative   39; Mismatches   73; Indels   52; Gaps   7;
```

QY

```
11 TTVEKFLINKDTGEV-----SELPKPRVTYTIQNGKEMSTIVSEDFI 54  
|||:::||| | || | |: :|:
```



Db 391 TTAK--MSSSTAELVOLPAKGVKMDOKVSGEPMHLETAFTKDSAMEQIVAEADVAM 447  
 QY 55 LPPYKELKGVQFDCMEISGFECKKDACGVINLSKDTIKVPFKKEEKE-----106  
 Db 448 NPIVEKAMEWEAEAGAAINPIVEADG-----ANNPIVEKAMSQIVAEADAINQAVD 501  
 QY 107 ---EENKPPFVDVSKKKDNFO--VNHSQLNESHKREKLOREHSHQSKSDSTKVATVATLDXN 161  
 Db 502 ANRQOTAPFGNDAAEDDDSEPSHS---FTLNPLPEKKEVWRKQATRSVSADQODXN 558  
 QY 162 N-----ISSKSTNNPNK 174  
 Db 559 SKTIAESSLOEISASQTENSPFO 582

RESULT 39  
 ORMSPI  
 microtubule-associated protein MAP1B - mouse  
 N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei  
 C:Species: Mus musculus (house mouse)  
 C>Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #ext\_change 01-Sep-2000  
 C/Accession: S07549, S44387, A33645  
 R/Mobile, M.; Lewis, S.A.; Cowan, N.J.  
 J. Cell Biol. 109, 3367-3376, 1989  
 A>Title: The microtubule binding domain of microtubule-associated protein MAP1B contains  
 A:Reference number: A33645, MUID:90094539, PMID:2480963  
 A/Accession: S07549  
 A/Molecule type: mRNA  
 A:Residues: 1-2464 <NOB>  
 A/Cross-references: EMBL:X51396; NID:952999; PIDN:CA53761.1; PID:953000  
 R/Sanchez, C.; Paducci, R.; Paciucci, R.; Zabala, J.C.; Avila, J.  
 Arch. Biochem. Biophys. 310, 428-432, 1994  
 A>Title: Binding of heat-shock protein 70 (hsp70) to tubulin.  
 A:Reference number: S44387, MUID:94234720, PMID:8179328  
 A/Accession: S44387  
 A/Molecule type: protein  
 A:Status: preliminary  
 A:Residues: 653-663, 'IC' <SAN>  
 C:Superfamily: microtubule-associated protein MAP1B  
 C/Keywords: microtubule binding; phosphoprotein; tandem repeat  
 F:589-786/Domain: microtubule binding #status experimental <MTB>  
 F:589-592, 639-642, 649-652, 655-658, 660-663, 668-671, 674-677, 679-682, 683-686, 687-690, 691-69  
 R-X/E/D-X)  
 F:1861-2064/Region: 17-residue repeats  
 F:116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: PH  
 F:447,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (Co  
 F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 10.4%; Score 93; DB 1; Length 2464;  
 Best local Similarity 27.8%; Pred. No. 1,2e+02;  
 Matches 49; Conservative 24; Mismatches 57; Indels 46; Gaps 11;

QY 14 KEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEKGVQFGMEI 73  
 Db 584 EKVLVKKDKPVKTESKP---SVT---EKEVSS---KEEQ--SPV-KAEVA-----EK 623  
 QY 74 SGEFGKDKAGYVNIKQPIKRVK-KTEEKEENKPPFVDSKKKDNQVNHQSQNS 132  
 Db 624 QATESKP-----KVTKDKVVKKEITKLEKEE--KPKKEVKKEDTTP--KDKK 672  
 QY 133 HKREDKQRE-----EHSQKSDSTKVATVATLDKNISKSSTNNPNK 174  
 Db 673 PRKEVKEIKKEIKKEERKELEKGVKKEKTPPLDKAKVKEKEKEKKEKPKK 728

RESULT 40  
 S66040  
 serine-type D-Ala-D-Ala carboxypeptidase (EC 3.4.16.4) dacc - Bacillus subtilis  
 N:Alternate names: penicillin-binding protein 5  
 C/Species: Bacillus subtilis  
 C>Date: 28-Oct-1996 #sequence revision 13-Mar-1997 #ext\_change 20-Jun-2000  
 C/Accession: S66040, I39830, A92275, A92307, B61335, D69612, A23307  
 R/Ogasawara, N.; Nakai, S.; Yoshikawa, H.

DNA Res. 1, 1-14, 1994  
 A>Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom  
 A:Reference number: S65967, MUID:96051385, PMID:7584024  
 A/Accession: S66040  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A:Residues: 1-443 <OGA>  
 A/Cross-references: EMBL:D26185; NID:9467326; PIDN:BA05246.1; PID:9467400  
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993  
 R/Todd, J.A.; Roberts, A.N.; Johnstone, K.; Piggett, P.J.; Winter, G.; Ellar, D.J.  
 J. Bacteriol. 167, 257-264, 1986  
 A>Title: Reduced heat resistance of mutant spores after cloning and mutagenesis of the B  
 A:Reference number: I39830, MUID:86250602, PMID:3087956  
 A/Accession: I39830  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A:Residues: 48-226, 'O', 228-443 <RES>  
 A/Cross-references: GB:M13766; NID:9142816; PIDN:AAA22375.1; PID:9142817  
 R/Waxman, D.J.; Strominger, J.L.  
 J. Biol. Chem. 255, 3964-3976, 1980  
 A>Title: Sequence of active site peptides from the penicillin-sensitive D-alanine carbox  
 A:Reference number: A92275, MUID:80182289, PMID:6768745  
 A/Accession: A92275  
 A/Molecule type: protein  
 A:Residues: 32-95, 'X', 97-98, 'XOX', 102 <MAX>  
 R/Waxman, D.J.; Strominger, J.L.  
 J. Biol. Chem. 255, 2067-2077, 1981  
 A>Title: Primary structure of the COOH-terminal membranous segment of a penicillin-sensi  
 A:Reference number: A92307, MUID:8117303, PMID:6780559  
 A/Accession: A92307  
 A/Molecule type: protein  
 A:Residues: 414-443 <MAX>  
 R/Yocum, R.R.; Waxman, D.J.; Rasmussen, J.R.; Strominger, J.L.  
 Proc. Natl. Acad. Sci. U.S.A. 76, 2730-2734, 1979  
 A>Title: Mechanism of penicillin action: penicillin and substrate bind covalently to the  
 A:Reference number: A61335, MUID:79233865, PMID:111240  
 A/Accession: B61335  
 A/Molecule type: protein  
 A:Residues: 55-68 <YOC>  
 R/Kunze, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
 C.; Bron, S.; Brouillette, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
 A.; Ehrlich, S.D.; Emerson, P.T.; Eutlan, K.D.; Erttington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulter, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Gallazzi, A.; Gall  
 lech, J.; Harwood, C.R.; Hensut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koester, P.; Koningsstein, G.; Krogh, A.; Kumano, M.; Kunita, K.; Lapidus, S.; Lardinois,  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Maeda, S.; Maue  
 Y, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror  
 akouch, M.; Tanakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K  
 A:Authors: Yoshikawa, H.F.; Zumestein, E.; Yoshikawa, H.; Danchin, A.  
 A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580, MUID:98044033, PMID:9384377  
 A/Accession: D69612  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A:Residues: 1-443 <KUN>  
 A/Cross-references: GB:D29104; GB:AL009126; NID:92632267; PIDN:CAB1786.1; PID:92632277  
 A:Experimental source: strain 168  
 C/Genetics:  
 A:Gene: dacc  
 A:Start codon: TTG  
 C:Superfamily: penicillin-binding protein 5  
 C/Keywords: cell wall synthesis; hydrolase; serine carboxypeptidase  
 F:67/Active site: Ser #status experimental

Query Match 10.3%; Score 92.5; DB 2; Length 443;  
 Best local Similarity 25.0%; Pred. No. 19;  
 Matches 36; Conservative 26; Mismatches 49; Indels 33; Gaps 6;

QY 6 DPA-RNTVKEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEK 64

Db 304 DYAFNPFMKEIYABD-----QVKGHK-TISVDKGEKEVGIVTKAFSLPVKNGE-EK 356  
OY 65 GYQFDGWEISGFEKKDAGYVNLKSDTFIKPVFKKIEKKKEENKPTFDVSKKKDNPOV 124  
Db 357 NYKAK-----VTLNKDNLTAPVKKGTKVGK-----LTAETGDEK 391  
OY 125 NHSQINESHKEDLOREHSOKSD 148  
Db 392 DYGLNSDLAGVDLVTKENVERKAN 415

Search completed: February 10, 2004, 10:58:37  
Job time : 14.5638 secs

***This Page Blank (uspto)***

DR PRINTS; PRO1228; EGGSHELL.  
KW Hypothetical protein; Signal; Transmembrane.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 385 HYPOTHETICAL GLYCINE-RICH PROTEIN  
FT TRANSMEM 165 185 POTENTIAL.  
FT DOMAIN 23 190 GUY-RICH.  
SQ SEQUENCE 385 AA; 36963 MW; 32AB25AE9B5073FE CRC64;

Query Match 11.7%; Score 99.5; DB 5; Length 385;  
Best Local Similarity 27.7%; Pred. No. 13;  
Matches 46; Conservative 26; Mismatches 51; Indels 43; Gaps 8;

OY 2 TVKEPLIND-----TGEVSELKPHRYVTITQ--NGKMSSTIVSEDP1 44  
DB 232 TLKEKAGQDIYVVCNEKAPFTAFETDDFCSLQENVHCITLIRINKEVAK--NEEDKK 289  
OY 45 LPVYKGELEKGYQPDGMEISGPGKDGAVINLSKOTFIKPVFKIEEKEENKPTD 104  
DB 290 EEBPKKEEKEKEEKEEKEE--EDEKDE-----EP--KKEEKEEKEEKE--E 330  
OY 105 VSKKKDPQVNHSQLNESHKEDLQREHSQKSDSTQVATATVLDK 150  
DB 331 VERKEE-----EKKOEPEPKKEEKEEKEEKEDEVEKSEKVBK 371

## RESULT 39

O813J9 PRELIMINARY; PRT; 4524 AA.  
AC O813J9;  
DT 01-MAR-2003 (TRENBLREL. 23, Created)  
DT 01-MAR-2003 (TRENBLREL. 23, last sequence update)  
DT 01-MAR-2003 (TRENBLREL. 23, last annotation update)  
DE Hypothetical protein.  
GN PF31325W.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Devlin K., Baker S., Davies P., Mungai K., Berriman M., Pain A.,  
RA Hall N., Bowman S., Churcher C., Quail M., Barrett B.,  
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22255708; PubMed=12368867;  
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
RA Mungai K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,  
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,  
RA Cronin A., Davies P., Davis R., Dear P., Dearden F., Doggett J.,  
RA Felwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
RA Harper D., Hauser H., Hornby T., Holroyd S., Horrocks P.,  
RA Humphrey S., Jagels K., James K.D., Johnson D., Kethorou A.,  
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,  
RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,  
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinovitch B.,  
RA Rajadream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,  
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,  
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,  
RA Sultoni J.E., Craig A., Newbold C., Barrett B.G.  
RT "Sequence of plasmodium falciparum chromosomes 1, 3-9 and 13.";  
RL Nature 419:527-531(2002).  
DR EMBL; AL929354; CAD51629.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 4524 AA; 525287 MW; 542206173C29567A CRC64;

Query Match 11.7%; Score 99.5; DB 5; Length 4524;  
Best Local Similarity 24.3%; Pred. No. 2e+02;  
Matches 44; Conservative 27; Mismatches 81; Indels 29; Gaps 7;

OY 1 TVKEPLIND-----TGEVSELKPHRYVTITQ--NGKMSSTIVSEDP 43

DB 3812 TNESEVITKDKIIMNDVPTQEBFDIKPEVESNLIESNNNNKKSKNNIYKORSM 3871  
OY 44 ILPVYKGELEKGYQPDGMEISGPGKDGAVINLSKOTFIKPVFKIEEKEENKPTD 103  
DB 3872 VHYINQV---YFPNSIRYSNFNLSIDLN-MINLANKN-SKITTDLISKCIHNSMS 3926  
OY 104 DVSKKDPQVNHSQLNESHKEDLQREHSQKSDSTQVATATVLDKNNISKSTTNNP 163  
DB 3927 SIIRKYSMKIKKSVIKKISMKNIIOKET-EKKQDVYREV-----KNNLKRSNDANKN 3979  
OY 164 K 164  
DB 3980 K 3980

## RESULT 40

P91488 PRELIMINARY; PRT; 211 AA.  
AC P91488;  
DT 01-MAY-1997 (TRENBLREL. 03, Created)  
DT 01-MAY-1997 (TRENBLREL. 03, last sequence update)  
DT 01-MAR-2003 (TRENBLREL. 23, last annotation update)  
DE Hypothetical 24.0 kDa protein.  
GN T23B3.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Briscot N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Briscot N2;  
RA Magg L., Le T.,  
RT "The sequence of C. elegans cosmid T23B3.";  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Briscot N2;  
RA Waterston R.,  
RT "Direct Submission.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U08309; AAB42334.1; -.  
DR WormPep; T23B3.5; CRI4016.  
KW Hypothetical protein.  
SQ SEQUENCE 211 AA; 23955 MW; 11B1614A87E5928 CRC64;

Query Match 11.7%; Score 99; DB 5; Length 211;  
Best Local Similarity 29.4%; Pred. No. 7.6;  
Matches 30; Conservative 17; Mismatches 51; Indels 4; Gaps 1;

OY 67 EGGKAGYVINSKOTFIKPVFKIEEKEENKPTFVSKKDPQVNH-----SQLNES 122  
DB 69 EGGKAGYVINSKOTFIKPVFKIEEKEENKPTFVSKKDPQVNH-----SQLNES 128  
OY 123 HRKEDLQREHSQKSDSTQVATATVLDKNNISKSTTNNP 164  
DB 129 EKKDDKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEKDEK 170

Search completed: February 10, 2004, 10:57:09  
Job time : 31.0366 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2004, 10:48:44 ; Search time 30.8071 Seconds  
(without alignments)  
1457.493 Million cell updates/sec

Title: US-10-067-385-8\_COPY\_600\_773

Perfect score: 897

Sequence: 1 KIVKQDFRNTTVKFFILNK.....ATVLDKNNISSKSTNNPNK 174

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lasting first 45 summaries

Database :

SPTREMBL.23.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_ricinet:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	897	100.0	2119	2 Q9AHT5	Q9AHT5 streptococc
2	897	100.0	2140	16 Q97RY6	Q97RY6 streptococc
3	894	99.7	2144	16 Q8DOP7	Q8DOP7 streptococc
4	893	99.6	2144	2 Q9S4M8	Q9S4M8 streptococc
5	119	13.3	775	16 Q8CPK8	Q8CPK8 streptococc
6	114.5	12.8	1038	13 Q90784	Q90784 gallus galli
7	114	12.7	609	5 Q812K8	Q812K8 plasmodium
8	112.5	12.5	346	5 Q9U0G0	Q9U0G0 plasmodium
9	112.5	12.5	354	5 Q2S995	Q2S995 plasmodium
10	112.5	12.5	354	5 Q81U55	Q81U55 plasmodium
11	112.5	12.5	361	5 Q9S4P5	Q9S4P5 plasmodium
12	112.5	12.5	379	5 Q9U6C4	Q9U6C4 plasmodium
13	112	12.5	829	5 Q815F3	Q815F3 plasmodium
14	111.5	12.4	379	5 Q2S706	Q2S706 plasmodium
15	110.5	12.3	379	5 Q2S705	Q2S705 plasmodium
16	110	12.3	380	5 Q26019	Q26019 plasmodium

17	109.5	12.2	3008	5 Q81436	Q81436 plasmodium
18	109	12.2	951	5 Q96229	Q96229 plasmodium
19	108.5	12.1	600	5 Q77355	Q77355 plasmodium
20	107.5	12.0	470	10 Q9FUK9	Q9FUK9 arabidopsis
21	107.5	12.0	948	5 Q9U4U6	Q9U4U6 plasmodium
22	106	11.8	3026	5 Q81LS9	Q81LS9 plasmodium
23	105.5	11.8	329	5 Q9NFV9	Q9NFV9 plasmodium
24	105.5	11.8	2081	10 Q9L498	Q9L498 arabidopsis
25	105	11.7	3127	5 Q81DA0	Q81DA0 plasmodium
26	104	11.6	540	10 Q94CS9	Q94CS9 arabidopsis
27	104	11.6	540	10 Q9S484	Q9S484 arabidopsis
28	103.5	11.5	328	5 Q44016	Q44016 dictyosteli
29	103	11.5	335	5 Q81226	Q81226 plasmodium
30	102.5	11.4	385	5 Q93424	Q93424 caenorhabdi
31	102.5	11.4	449	5 Q81HW3	Q81HW3 caenorhabdi
32	102	11.4	622	16 Q98Q41	Q98Q41 mycoplasma
33	102	11.4	1419	5 Q9UBD4	Q9UBD4 plasmodium
34	102	11.4	1612	5 Q81FNO	Q81FNO plasmodium
35	101.5	11.3	312	16 Q9PPL5	Q9PPL5 campylobact
36	101.5	11.3	382	5 Q9V7J0	Q9V7J0 drosophila
37	101.5	11.3	535	5 Q17595	Q17595 caenorhabdi
38	101.5	11.3	556	5 Q9V719	Q9V719 drosophila
39	101.5	11.3	785	5 Q9G082	Q9G082 drosophila
40	101.5	11.3	1377	5 Q81586	Q81586 plasmodium
41	101.5	11.3	1434	5 Q81J77	Q81J77 plasmodium
42	101	11.3	700	3 Q07457	Q07457 saccharomyc
43	101	11.3	1490	5 Q81416	Q81416 plasmodium
44	100.5	11.2	354	2 Q93CL2	Q93CL2 mycoplasma
45	100.5	11.2	1373	5 Q81353	Q81353 plasmodium

## ALIGNMENTS

### RESULT 1

Q9AHT5 PRELIMINARY: PRT: 2119 AA.  
AC Q9AHT5;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Serine protease (Fragment).  
GN PRTA.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N4;  
RX MEDLINE=21116976; PubMed=11179332;  
RA Witzemann T.M., Heinriche J.H., Adamou J.E., Erwin A.T., Kunsch C., Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E., Gayle A., Brewah Y.A., Walsh W., Barren P., Lachigra R., Hanson M., Langermann S., Johnson S., Koenig S.;  
RT "Use of a Whole Genome Approach to Identify Vaccine Molecules Affording Protection against Streptococcus pneumoniae Infection.";  
RT Infect. Immun. 69:1593-1598(2001).  
-- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY AN AMIDE BOND (BY SIMILARITY).  
EMBL: AF291699; AAK19159.1; --  
DR HSRP; P00782; 2SPT.  
DR MEROPS; S08.064; --  
DR InterPro; IPR001899; Gram\_pos\_anchor.  
DR InterPro; IPR006192; LPXTG.  
DR InterPro; IPR003137; PA.  
DR InterPro; IPR000209; Peptidase\_S8.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
DR Pfam; PF02225; PA; 1.  
DR Pfam; PF00082; Peptidase\_S8; 2.  
DR PRINTS; PR00723; SUBTILISIN.  
DR TIGRPFMs; TIGR01167; LPXTG\_anchor; 1.

DR PROSITE: PS0847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE: PS0840; PA; 1.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 1.  
 KW Cell wall; Peptidoglycan-anchor; Protease.  
 FT NON TER 1  
 SQ SEQUENCE 2119 AA; 238226 MW; 517F9B7F6B960A6A CRC64;  
 Query Match 100.0%; Score 897; DB 2; Length 2119;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-54;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KIVVDFPANTYVKEFIINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFIIIPVYKG 60  
 DB 1912 KIVVDFPANTYVKEFIINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFIIIPVYKG 1971  
 QY 61 ELEKGYQFDGWEISGFEKGKDGAVYINLSKDTFIRPVFKIIEKKEENKPTFDVSKKD 120  
 DB 1972 ELEKGYQFDGWEISGFEKGKDGAVYINLSKDTFIRPVFKIIEKKEENKPTFDVSKKD 2031  
 QY 121 NPQVNSQINSHRKEDLQREHSQKSDSTKDTATVLDKNNISSKSTNNPNK 174  
 DB 2032 NPQVNSQINSHRKEDLQREHSQKSDSTKDTATVLDKNNISSKSTNNPNK 2085  
 RESULT 2  
 Q97RY6 PRELIMINARY; PRT; 2140 AA.  
 ID Q97RY6;  
 AC Q97RY6;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Serine protease, subtilase family.  
 GN SP0641.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,  
 Durkin A.S., Gattam M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D.,  
 Holtapple E., Khouiri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
 McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus  
 pneumoniae";  
 RT Science 293:498-506(2001).  
 RL EMBL: AE007373; AAK74791.1; -.  
 DR MEROPS: S08\_064; -.  
 DR TIGR: SP0641; -.  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR InterPro: IPR006192; LPXTG.  
 DR InterPro: IPR003137; PA.  
 DR InterPro: IPR000209; Peptidase\_S8.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam: PF02225; PA; 1.  
 DR Pfam: PF00082; Peptidase\_S8; 2.  
 DR PRINTS: PR00723; SUBTILISIN.  
 DR TIGRPFMS: TIGR01167; LPXTG\_anchor; 1.  
 DR PROSITE: PS0847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE: PS0840; PA; 1.  
 DR PROSITE: PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE: PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 1.  
 KW Protease; Complete proteome.

SQ SEQUENCE 2140 AA; 240426 MW; FA44AD8E2938B334 CRC64;  
 Query Match 100.0%; Score 897; DB 16; Length 2140;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-54;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KIVVDFPANTYVKEFIINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFIIIPVYKG 60  
 DB 1933 KIVVDFPANTYVKEFIINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFIIIPVYKG 1992.  
 QY 61 ELEKGYQFDGWEISGFEKGKDGAVYINLSKDTFIRPVFKIIEKKEENKPTFDVSKKD 120  
 DB 1993 ELEKGYQFDGWEISGFEKGKDGAVYINLSKDTFIRPVFKIIEKKEENKPTFDVSKKD 2052  
 QY 121 NPQVNSQINSHRKEDLQREHSQKSDSTKDTATVLDKNNISSKSTNNPNK 174  
 DB 2053 NPQVNSQINSHRKEDLQREHSQKSDSTKDTATVLDKNNISSKSTNNPNK 2106  
 RESULT 3  
 Q8DOP7 PRELIMINARY; PRT; 2144 AA.  
 ID Q8DOP7;  
 AC Q8DOP7;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Cell wall-associated serine proteinase Prta (EC 3.4.21.-).  
 GN PRta OR SP00561.  
 GN Streptococcus pneumoniae (strain ATCC BAA-235 / R6).  
 OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=171101;  
 RN [1]  
 RA SEQUENCE FROM N.A.  
 RX MEDLINE=21429245; PubMed=11544234;  
 RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,  
 Dehoff B.S., Batrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,  
 Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,  
 LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,  
 McAhren S.M., McHenry M., Mcleaster K., Mundy C.W., Nicae T.I.,  
 Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,  
 Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,  
 Zook C.A., Baltz R.H., Tskunas S.R., Roestek P.R. Jr., Skatrud P.L.,  
 RA Glass J.I.;  
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6";  
 RL J. Bacteriol. 183:5709-5717(2001).  
 DR EMBL: AE008434; AAK9365.1; -.  
 KW Hydrolyase; Complete proteome.  
 SQ SEQUENCE 2144 AA; 240436 MW; 8C1B481DBC503A0C CRC64;  
 Query Match 99.7%; Score 894; DB 16; Length 2144;  
 Best Local Similarity 99.4%; Pred. No. 1.3e-53;  
 Matches 173; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KIVVDFPANTYVKEFIINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFIIIPVYKG 60  
 DB 1937 KIVVDFPANTYVKEFIINKDGEVSELKPHRVTVTIQNGKMSSTIVSEEDFIIIPVYKG 1996  
 QY 61 ELEKGYQFDGWEISGFEKGKDGAVYINLSKDTFIRPVFKIIEKKEENKPTFDVSKKD 120  
 DB 1997 ELEKGYQFDGWEISGFEKGKDGAVYINLSKDTFIRPVFKIIEKKEENKPTFDVSKKD 2056  
 QY 121 NPQVNSQINSHRKEDLQREHSQKSDSTKDTATVLDKNNISSKSTNNPNK 174  
 DB 2057 NPQVNSQINSHRKEDLQREHSQKSDSTKDTATVLDKNNISSKSTNNPNK 2110  
 RESULT 4  
 Q9S4M8 PRELIMINARY; PRT; 2144 AA.  
 ID Q9S4M8;  
 AC Q9S4M8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)



01-MAR-2003 (TRENBLREL. 23, Last annotation update)  
 DE Cell wall-associated serine proteinase Prta precursor.  
 GN PRTA.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 CC Streptococcus.  
 NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3.B;  
 RA Behe G., ten Thoren E., Bongers R.J.M., Heinz H.-P., Zysk G.,  
 RT "Cloning and sequencing of a novel surface protease of Streptococcus  
 pneumoniae."  
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE CELL WALL PEPTIDOGLYCAN BY  
 AN AMIDE BOND (BY SIMILARITY).  
 DR EMBL; AF127143; AAD48399.1; -.  
 DR HSSP; P00782; 2SPT.  
 DR MEROPS; S08.064; -.  
 DR InterPro; IPR001699; Gram\_pos\_anchor.  
 DR InterPro; IPR006192; LPTG.  
 DR InterPro; IPR003137; PA.  
 DR InterPro; IPR000209; Peptidase\_S8.  
 DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR Pfam; PF02225; PA\_1.  
 DR Pfam; PF00082; Peptidase\_S8; 2.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR TIGRFAMs; TIGR01167; LPTG\_anchor; 1.  
 DR PROSITE; PS00847; GRAM\_POS\_ANCHORING; 1.  
 DR PROSITE; PS00840; PA\_1.  
 DR PROSITE; PS00137; SUBTILASE\_HIS; 1.  
 DR PROSITE; PS00138; SUBTILASE\_SER; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 1.  
 DR Cell wall, peptidoglycan-anchor; Signal.  
 KW SIGNAL.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 2144 CELL WALL-ASSOCIATED SERINE PROTEINASE  
 FT SEQUENCE 2144 AA; 240724 MW; 2052511470741331 CRC64;  
 SQ  
 Query Match 99.6%; Score 893; DB 2; Length 2144;  
 Best Local Similarity 98.9%; Pred. No. 1.5e-53;  
 Matches 172; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KIVVDFARNTTVEKFIINKDTGEVSELRPHRVTVTIQNGKMSSTIVSEBDFILPYK 60  
 DB 1937 KIVVDFARNTTVEKFIINKDTGEVSELRPHRVTVTIQNGKMSSTIVSEBDFILPYK 1996  
 QY 61 ELEKGYQDGMWISFEGEKKAGYVINSKOTFIKPVFKIIEKKEENKPTFDVSKKD 120  
 DB 1997 ELEKGYQDGMWISFEGEKKAGYVINSKOTFIKPVFKIIEKKEENKPTFDVSKKD 2056  
 QY 121 NPQVNSQLNESHKEDLOREHESQKSDTKDVTATVLDKNNISKSTNNPNK 174  
 DB 2057 NPQVNSQLNESHKEDLOREHESQKSDTKDVTATVLDKNNISKSTNNPNK 2110  
 RESULT 5  
 OSCP88 PRELIMINARY; PRT; 775 AA.  
 ID OSCP88  
 AC OSCP88  
 DT 01-MAR-2003 (TRENBLREL. 23, Created)  
 DT 01-MAR-2003 (TRENBLREL. 23, Last sequence update)  
 DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)  
 DE Penicillin-binding protein 1.  
 GN SE0856  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 NCBI\_TaxID=1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,

Chen Z., Wen Y.;  
 RA Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AE016746; AA004453.1; -.  
 KW Complete Proteome.  
 SQ SEQUENCE 775 AA; 86354 MW; 4A00563A7BB8777C CRC64;  
 Query Match 13.3%; Score 119; DB 16; Length 775;  
 Best Local Similarity 27.0%; Pred. No. 2;  
 Matches 47; Conservative 24; Mismatches 57; Indels 46; Gaps 9;  
 QY 20 KDTGEVSELRPHRVTVTIQNGKMSSTIVSEBDFILPYK-----GELEKGYQFDGM-- 71  
 DB 614 EDSVNAQSLKP-----ITIGNGKQIKQSGVSGTKVLPHSKYMLMTDGLTWP-DWTGWTK 668  
 QY 72 -EISGFE-----GKQAGYIN--LSKDTFIKPVFKIIEKKEENKPTFDVSKKD 117  
 DB 669 EDVLAFFEDLTKIKVSTKNGFVTNQSISKQIIR-----NKDKIEVLSAED 715  
 QY 118 KKDNPQVNSQLNESHKEDLOREHESQKSDTKDVTATVLDKNNISKSTNN 171  
 DB 716 TDDDEKTDSDSKSKDQADBDHSNTSSITKN-----DKSNADSKNDSDD 763  
 RESULT 6  
 O90784 PRELIMINARY; PRT; 1038 AA.  
 ID O90784  
 AC O90784  
 DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)  
 DE Clastrin.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Cole G.J.;  
 RL Submitted (OCT-1992) to the EMBL/Genbank/DBJ databases.  
 RP SEQUENCE OF 1-451 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=94157526; PubMed=7906711;  
 RA Burg M.A., Cole G.J.;  
 RT "Clustrin, an antiadhesive neural keratan sulfate proteoglycan, is  
 RT structurally related to MAP1B."  
 RL J. Neurobiol. 25:1-32(1994).  
 DR EMBL; X67778; CAA47988.1; -.  
 SQ SEQUENCE 1038 AA; 117112 MW; 213D694A5B510927 CRC64;  
 Query Match 12.8%; Score 114.5; DB 13; Length 1038;  
 Best Local Similarity 23.5%; Pred. No. 5.7;  
 Matches 47; Conservative 29; Mismatches 65; Indels 59; Gaps 6;  
 QY 2 IVVDFARN---TVKEFIINKDTGEVSELRPHRVTVTIQ-----GELEKGYQFDGM-- 39  
 DB 514 VTQKDLTGNIASPAVKQAKLQRTDSKESLPAAKTTTKQCKRNLKKKHWISLQSLVQ 573  
 QY 40 -----GKMSSTIVSEBDFILPYKGELEKGYQFDGMETSGFSGK 79  
 DB 574 LEKPKLESKEKTPYKKAQKAVKPEKTVAKRDV-----TTKEQLGSESEKQAS 625  
 QY 80 KDAGVINTLSKDTFIKPVFK-KIEKKEENKPTFDVSKKDNQVNSQLNESHKEDL 138  
 DB 626 EKQDVKPVTVTEKSVKKEVAKPEKDKBRKPKKVEYSKKEKPLI---KKEKKKKEDI 682  
 QY 139 QRE-----EHSQKSDTKDV 153  
 DB 683 KKEVKEVKEKKEKKEKKEV 702



Db	309 NNN 311
RESULT 10	
Q81J55	
ID	Q81J55
AC	Q81J55
NC	Q81J55
DT	01-MAR-2003 (TREMBLrel. 23, Created)
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE	Merozoite surface protein 3.
GN	PF10_0345.
OS	Plasmodium falciparum (isolate 3D7).
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_Taxid=36929;
LN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=3D7;
RC	MEDLINE=22255705; Pubmed=12368864;
RA	Gardner M.J., Hall N., Fung E., White O., Beriman M., Hyman R.W.,
RA	Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA	Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA	Chan W.-S., Nene V., Shallow S.J., Sub B., Peterson J., Angioli S.,
RA	Petrea M., Allen J., Selengut J., Haft D., Mather M.W., Waideya A.B.,
RA	Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA	McCarthy G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA	Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA	Fraser C.M., Barrall B.;
RT	"Genome sequence of the human malaria parasite Plasmodium
RT	falciparum.";
RL	Nature 419:498-511(2002).
RL	EMBL; AE014834; AA035542.1; -
Q8	SEQUENCE 354 AA; 40119 MW; 3A7256152F48B527 CRC64;

	Query Match	Similarity	22.4%	Pred. No.	2.5;	Mismatches	62;	Indels	45;	Gaps	7
	Matches	41;	Conservative	35;							
Dy	10	NTTVEKFLIN-KDTGEVSELKPRHTVTYIIONGKMSSTIVSEDEPILPYKKLELKGYPF	68								
Dd	153	STKTREYAEKAKNAAYEKAKMAYQKANKQAVLKAKEAS-----YDYIL	194								
Dy	69	DGMELISGF-ECKKDAG-----YVINLSKDFIKVPFKIEKKKE-----E	108								
Dd	195	-GWETGGGVPPHKKKEENMLSHLYVSXCKKENISKNDVDLDERKEEAETEELKEKNKE	253								
Dy	109	NKPFTDVSKKKDNPQVNHSQNLNEHRKDDLQREHSQKSDSKDYATATYALVDNNINSKST	168								
Dd	254	EETSEISEDEEEEEEKEENDKKKEQEKEQSENNDDKKDWEA-----QNLSKNO	308								
Dy	169	TNN 171									
Dd	309	NNN 311									

-RESULT 11

ID	OSPT5	PRELIMINARY:	PRT:	361 AA.
AC	O9SP15			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)			
DE	Merozoite surface protein 3 (Fragment).			
OS	Plasmodium falciparum.			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5633;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FVO;			
RA	Hisaeda H., Saul A., Long C.A., Miller L.H., Stowers A.W.;			
RT	"Merozoite Surface Protein 3 and Protection Against Malaria in Anous			
RL	Monkeys.";			
	Submitted (JUL-2001) to the EMBL/GenBank/DBSJ databases.			

DR	EMBL; AY044180; AAKK94780.1; -.
FT	NON_TER 1
FT	NON_TER 361
SQ	SEQUENCE 361 AA; 41163 MW; 6127A3041587BA74 CRC64;

```

Query Match      12.5%; Score 112.5; DB 5; length 361;
Best Local Similarity 22.4%; Pred. No. 2.5;
Matches 43; Conservative 33; Mismatches 55; Indels 61; Gaps 8

QY      21 DTGEVSELKPHRYVTVLTLONGKMSSTVSSEDFILIPVKKELEKYQ----- 67
Db       147 ETGG-----:::|:::|:::|:::|:::|:::|:::|:::|:::|:::| 187
          RNSHNPFYTXTKE-----YAGKVKKDYEEAKNAYOKANQAV
QY      68 -----FD--GMEISGF--EGKKOAG-----YVILSKDTFLKPVPKIAEKKKEEN 109
Db       188 LKAKEASSYDILLGMERGGGVPEHKKEENMTLSHYSSKDKNENISKENDVDLDE-KEBEA 246
          :|||:|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      110 KPTFDSVKKKDNPVNHSQNLNESHRKEDLOREESHQSQSDSTDYATVATVLDKN----- 161
Db       247 EETEEBELAEKNEBETESEIIEDEDEEEEEEEKKEENDKKKEQCKEQESNNENNDDKKWE 306
          ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      162 --NISKSNTNN 171
Db       307 AQNLTISKNNNN 318

```

RESULT 12			
09UC4			
09UC4	PRELIMINARY;	PT;	379 AA.
AC	Q9UC4;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, last annotation update)		
DE	Polymorphic antigen.		
GN	MSP-3.		
OS	Plasmodium falciparum.		
OC	Eukaryotes; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=5833;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FC01/HN;		
RA	Li X.B., Yu X.B., Luo S.H., Shan Z.X., Fang J.M., Xu J.;		
RT	"Sequence of Plasmodium falciparum secreted polymorphic antigen		
RL	gene.";		
SR	Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF188190; AF04099.1; -. C152A54E1P9D5F25 CRC64;		
Q0	SEQUENCE 379 AA; 43316 MW; C152A54E1P9D5F25 CRC64;		

Query Match	12.5%;	Score 112.5;	DB 5;	Length 379;
Best Local Similarity	23.6%;	Pred. No. 2.7;		
Matches 41;	Conservative 35;	Mismatches 59;	Indels 39;	Gaps 8

[illegible]

RESULT 13	
0815F3	
ID 0815F3	PRELIMINARY; PRT; 829 AA.
AC 0815F3	
DT 01-MAR-2003	(REMBLrel. 23, Created)
DT 01-MAR-2003	(REMBLrel. 23, Last sequence update)
DT 01-MAR-2003	(REMBLrel. 23, Last annotation update)
DE Hypothetical protein.	

CN PFL1275C.  
 OS Plasmodium falciparum (isolate 3D7).  
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner W.J., Hall N., Fung E., White O., Berrieman M., Hyman R.W., K.,  
 RA Carlton J.M., Pahn A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan W.-S., Nene V., Shallow S.J., Suh B., Peterson J., Anguilo S.,  
 RA Perlea M., Allen J., Selengut J., Hart D., Mather M.W., Valdiva A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraumholz M.J., Roos D.S., Ralph S.A.,  
 RA McCadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Fraser C.M., Barrrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RL falciparum.";  
 RL Nature 419:498-511(2002).  
 DR EMBL; AE014848; AAN36341.1; .  
 KW Hypothetical protein.  
 SQ SEQUENCE 829 AA; 98816 MW; EF2675E301B2CE93 CRC64;  
  
 Query Match 12.5%; Score 112; DB 5; Length 829;  
 Best Local Similarity 24.0%; Pred. No. 6.7;  
 Matches 44; Conservative 39; Mismatches 58; Indels 42; Gaps 9.  
  
 QY 8 ARNTVFKEFLINKDNGEVSSELPKPHVVTYIQNGKMSSTIVSEDFILPVYKGELEKGYQ 67  
 DB 487 AATTISNDISLNTKTKKI--IOP--LEVIYIKNGID--KTIWFERIVLDVPHPIKIKYH 540  
 QY 68 FDGEWISGEFGKQAGYVYINISKDTFIKPFKEIEKKEENKPT-----F 113  
 DB 541 LN-----KKETLTFNPSLN-----FREIKKRRDKKKGTHINNKDAEEMLYKY 583  
 QY 114 DYSKKKDN-PQVNHQQLNESH-RKEDDLQREHSQSDSKYVATVLDKNNISKSTNN 171  
 DB 584 KIKKKKKFNEENNTLNDNSIKKENNKLVHE--DNSLKORQIILNDKVIETHTKIYDN 640  
 QY 172 PNK 174  
 DB 641 QKK 643  
  
 RESULT 14  
 ID Q25706 PRELIMINARY; PRT: 379 AA.  
 AC Q25706;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)  
 DE Polymorphic antigen.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cs12;  
 RX MEDLINE=98156743; PubMed=9497029;  
 RA McColl D.J., Anders R.F.;  
 RT "Conservation of structural motifs and antigenic diversity in the  
 RT Plasmodium falciparum merozoite surface protein-3 (MSP-3).";  
 RL Mol. Biochem. Parasitol. 90:21-31(1997).  
 RL EMBL; U08852; AAC47832.1; .  
 SQ SEQUENCE 379 AA; 43302 MW; ABF9D54E1ED91A24 CRC64;  
  
 Query Match 12.4%; Score 111.5; DB 5; Length 379;  
 Best Local Similarity 23.6%; Pred. No. 3.1;  
 Matches 41; Conservative 35; Mismatches 59; Indels 39; Gaps 8.  
  
 QY 29 KHRVTVTIQNGKMSSTIVSEDF-----ILPVYKGELEKGYQFD-GWEISGF- 76

```

Db      171 KPSRLNFSRTKTEKAEV--EKDYERAKNAVQKANQAVLKAKKASSVYILGMEVGGCV 228
Qy      77 -EGKDGAG-----YVINSKDTFIKPVFKLIEKKKEENKPFVDVSKKKNQOVNHSQLN 130
Db      229 PEHKKEENMLSHLYVSSKDKENISKENDVDVDE-KEEEAEETEEBELEKKEEETEESEIS 287
Qy      131 -----BSHRKEDIQREHSQKSDSTKQVTAFTVLDKNNISSKSTTNN 171
Db      288 EDEEEEEEKKEENDKKKEQEKESQENENNDDQKMEA-----QNLISKNNNN 336

RESULT 15
ID      Q25705      PRELIMINARY;      PRT;      379 AA.
AC      Q25705;
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT      01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE      Polymorphic antigen.
OS      Plasmodium falciparum.
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_TaxID=5833;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=KL;
RX      MEDLINE=98156743; PubMed=9497029;
RA      MCColl D.J., Anders R.F.;
RT      "Conservation of structural motifs and antigenic diversity in the
RT      Plasmodium falciparum merozoite surface protein-3 (MSP-3).";
RL      Mol. Biochem. Parasitol. 90:21-31(1997).
DR      EMBL, U08851, AAC47831.1; -.
SQ      SEQUENCE 379 AA; 43344 MW; DCFATFI06887CGAAO CRC64;

Query Match      12.3%; Score 110.5; DB 5; Length 379;
Best Local Similarity 22.1%; Pred. No. 3.7;
Matches 43; Conservative 35; Mismatches 48; Indels 69; Gaps 9

Qy      21 DTGEVSELKPRHRYVTYIQTNGKEMSTIYSEDPFILPYRGELKQVQ----- 67
Db      167 ETGE-----RNSRNNFYTTKTE-----YAGKVEKDYERAKNAVQKANQAV 207
Qy      68 -----FD---GMEISGF--EGKDGAG-----YVINSKDTFIKPVFKLIEKKKEEN 109
Db      208 LKAKASSYDYLIGMEFGGVEPEHKKEENMLSHLYVSSKDKENISKENDVDVDE-KEEEA 266
Qy      110 KTFPVSKKKNQOVNHSQLN-----BSHRKEDIQREHSQKSDSTKQVTAFT 156
Db      267 EETEEBELEKKEEETEESEISDEDEBEBEKKEEENEKKQEKESQENENNDDQKMEA- 325
Qy      157 VLDKNNISSKSTTNN 171
Db      326 ----QNLISKNNNN 336

RESULT 16
ID      Q26019      PRELIMINARY;      PRT;      380 AA.
AC      Q26019;
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT      01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE      Polymorphic antigen precursor.
OS      Plasmodium falciparum.
OC      Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX      NCBI_TaxID=5833;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=PC27;
RX      MEDLINE=951598774; PubMed=7891748;
RA      MCColl D.J., Silva A., Foley M., Kun J.F., Favaloro J.M.,
RA      Thompson J.K., Marshall V.M., Coppel R.L., Kemp D.J., Anders R.F.;
RT      "Molecular variation in a novel polymorphic antigen associated with
RT      Plasmodium falciparum merozoites."

```

[illegible]

QY 5 KDFARTVKEFLINKDGEVSELKPHRYL-VTIQNGKEMSSITVEEDFILPVYKGELE 63  
 DB 115 KODNNNNNGTKOIEKNGKINKSDL-HRONELNLSGK-----NEQOI-----NKK 159  
 QY 64 KGYQDFGWEI--SGFGKKDAGVNLKSDTFIKPVFKKIEEKE-----EENKP 111  
 DB 160 KGGQ-----DISNNAENKKD-----VKEGVKEIEEKKKEKISDDHKVEENK 203  
 QY 112 TRD---VSKKDNPPVNSQANESHKRDQOR-EHSHQKSDSTKVATVLDKNNISK 166  
 DB 204 SDHKVEENKSDDHVEENKSDHKIEVKVEEHEDEE-----DKKKKSE 254  
 QY 167 STNNPNK 174  
 DB 255 NKKKDKNK 262

## RESULT 19

ID 077355 PRELIMINARY; PRT; 600 AA.  
 AC 077355;  
 DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Hypothetical 71.7 kDa protein.  
 GN PFC0465C, MAL34.20.  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=99376085; PubMed=10448855;  
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,  
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltham T.,  
 RA Gentles S., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,  
 RA Horrocks P., Jagels K., Jassal B., Kyes S., McLean J., Moule S.,  
 RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,  
 RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,  
 RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;  
 RA "The complete nucleotide sequence of chromosome 3 of Plasmodium  
 RT falciparum."  
 RL Nature 400:532-538(1999).  
 DR EMBL: AL008970; CA15610.2; --  
 DR InterPro: IPR002483; PFI.  
 DR Pfam: PF01480; PFI; 1.  
 DR SMART: SM00311; PFI; 1.  
 DR KW Hypothetical protein.  
 SQ SEQUENCE 600 AA; 71663 MW; 57EAB42565CAD64C CRC64;

Query Match 12.1%; Score 108.5; DB 5; Length 600;  
 Best Local Similarity 29.3%; Pred. No. 8.3;  
 Matches 54; Conservative 22; Mismatches 47; Indels 61; Gaps 12;

QY 47 IVESEDFILPVY-----KGELEKGYQDFGWEISGFEK-----DAGVINLSKDTFIKPY 97  
 DB 60 ILFEDFDILVEYICISQKSKK-----DGEEDRYLNAKKLNTLGTGNKKSDFILEL 116  
 QY 98 FKXI--EKKKE-----ENKPTFVSK-KKDNPPVNSQALNE-----SHRK 135  
 DB 117 LELLNEKKKEKHEIADTLNENK-TNDIKYKNNENINENVENNDISNKKDEHSHQN 175  
 QY 136 E-----DLQREH-----SQKSDSTK---DVTATVLDKNNISKSTTN 170  
 DB 176 EHNINNVNLKKEKEYTDIQDRKKRKHKSLSQKSDSYKRPFNKRTKSIER-SLSNRYDE 234  
 QY 171 NPKK 174  
 DB 235 KTKK 238

RESULT 20  
 Q9FJK9

ID Q9FJK9 PRELIMINARY; PRT; 470 AA.  
 AC Q9FJK9;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE GB|AF20218.1.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustoid II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=99087489; PubMed=9872454;  
 RA Nakamura Y., Sato S., Asamizu E., Kaneo T., Kotani H., Miyajima N.,  
 RA Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
 RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
 RT physically assigned P1 and TAC clones."  
 RL DNA Res. 5:297-308(1998).  
 DR EMBL: AB015468; BAB10694.1; --  
 SQ SEQUENCE 470 AA; 53758 MW; 6D686CE72B35AC54 CRC64;

Query Match 12.0%; Score 107.5; DB 10; Length 470;  
 Best Local Similarity 20.1%; Pred. No. 7.5;  
 Matches 36; Conservative 32; Mismatches 66; Indels 45; Gaps 5;

QY 29 KPHRYVTIQNGKEMSSITVSEEDFILPVYKGELEKGYQDFGWEISGFE-----GKK 80  
 DB 82 RNRVTDVTVQNNNSGSK-----YVDDLARIIRYDE-EATGSGSAGRIDHPNOK 129  
 QY 81 DAGVINLSKDTFIKPVFKKIEEKEENKPFEDVSKKDN----- 121  
 DB 130 NGITEKAPENSPLETSRVDNKRINNKNPFAKSEENAVSRVSGADHKRAEWCK 189  
 QY 122 POWNHQLNE-----SHKEDLQREHSHQKSDSTKVATVLDKNNISKSTNNPNK 174  
 DB 190 PHENDDQVQTSARSHKHKNVTSKPKRDEGVKTKTEAKXKDNKKEKKEKTSINK 248

## RESULT 21

ID Q9U4U6 PRELIMINARY; PRT; 948 AA.  
 AC Q9U4U6;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Ornithine decarboxylase.  
 GN ODC.  
 OS Plasmodium falciparum.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FCUPL/RSA;  
 RA Birholtz L., Joubert F., Neitz A.W.H., Louw A.I.;  
 RT "Molecular characterization of Plasmodium falciparum ornithine  
 RT decarboxylase cDNA obtained by RACE."  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF139900; AAF14518.1; --  
 DR HSSP: P07805; 1F3T.  
 DR InterPro: IPR000183; Decarboxylase2.  
 DR Pfam: PF02784; Orn\_Arg\_dec\_N; 1.  
 DR Pfam: PF00278; Orn\_DAP\_Arg\_dec; 1.  
 DR PRINTS: PR01179; ODADECARBLASE.  
 SQ SEQUENCE 948 AA; 110350 MW; 43P103DB83F12835 CRC64;

Query Match 12.0%; Score 107.5; DB 5; Length 948;  
 Best Local Similarity 22.0%; Pred. No. 16;  
 Matches 54; Conservative 32; Mismatches 71; Indels 89; Gaps 11;

QY 4 VKDFARTVKEFLINKDGEVSEL-----KPHRYVTIQNGKEMSSITVSEEDFIL 54

```

Db 547 VEDMSNNMGFNYYIINLGGGYDEELEYDPAKKDKHXYCTLSLQIKDKIQFLNEEFTL 606
Qy 55 LPVY-----KGELEKGYQFDGWEISGFEGRKD-----AGYINISKOTFIPVPEK 100
Db 607 KTKYGVYFEFKISLAINNSIDHY-----FSHMDNLRVICEPGRYVAASSTLAVXIGRR 662
Qy 101 IEKKEEENKPTF-----DVSKKDNPOVNSHREKED----- 137
Db 663 -----RPTFGIMLKDLKAHYDPLNFAQGEKKQDEPKINNNNNNDNNNDNN 713
Qy 138 LOREESQK-----SDST-----KDVATVLDK--NNIS-SKS 167
Db 724 NNNNNNNQGGGGINMDLIITSTNDSTYKKNDSHSSGYIQVGSCTIRKEDDNKIMTH 773
Qy 168 TTNPN 173
Db 774 TNNPN 779

```

## RESULT 22

```

Q8ILS9 PRELIMINARY; PRT; 3026 AA.
ID 08ILS9
AC 08ILS9;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN Pf14_0165.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung B., White O., Bertrman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eichen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chen W.-S., Nene V., Shalim S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Frazer C.M., Barrrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511(2002).
DR EMBL: AE014818; AAN36777.1; -.
KW Hypothetical protein.
SQ SEQUENCE 3026 AA; 357633 MW; 9ECABD915C3C25CE CRC64;

```

Query Match 11.8%; Score 106; DB 5; Length 3026;  
 Best Local Similarity 22.7%; Pred. No. 69;  
 Matches 44; Conservative 34; Mismatches 86; Indels 30; Gaps 8;

```

Qy 1 KIVVDPAFNTT-----VKEFIINKDTGEVSELKPHRVTVTTONGKMSSTIVSEDF 53
Db 146 KVLKESKDIQDYNLKDIDETEKRTMKSSYKQYNNMNCCTKNSSFNVTDN--- 202
Qy 54 ILPVKGELEKGYQFDGWEISGFEGRKDAGYINISKOTFIPVPEKIEKKEEENKPTF 113
Db 203 ---ICHGN-EKYKMSDNKQICIIIRKQOLIIDEIC--TMVKNANKKIKNOVEEYKNRV 256
Qy 114 DVSKKD---NPVNSHLSNREKEDLOREESQKS--DSTKDVATVLDKNISKST 168
Db 257 SVINRKNQNTQNSDINNTO-NILHNEDEIEYKLANENDIHTVKTITKEYSSNSGSSSD 315
Qy 169 T-----NNPNK 174
Db 316 TLLSYESVNNKKNK 329

```

```

RESULT 23
Q9NFV9 PRELIMINARY; PRT; 329 AA.
ID Q9NFV9
AC Q9NFV9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Merozoite surface protein 3 (Fragment).
GN MSP3.
OS Plasmodium falciparum (isolate 7G8).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=57266;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7G8;
RX MEDLINE=20416497; PubMed=10960178;
RA Okenu D.M.N., Thomas A.W., Conway D.J.;
RT "Allelic lineages of the merozoite surface protein 3 gene in
RT Plasmodium reichenowi and Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 109:185-188(2000).
DR EMBL: AJ252287; CAB85901.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 329 AA; 36916 MW; CSB045DBSE21A159 CRC64;

```

Query Match 11.8%; Score 105.5; DB 5; Length 329;  
 Best Local Similarity 24.2%; Pred. No. 7;  
 Matches 43; Conservative 31; Mismatches 63; Indels 41; Gaps 8;

```

Qy 10 NTWVEFLIN-KDTGEVSELKPHRVTVTTONGKMSSTIVSEDFILPVKGELEKGYQF 68
Db 143 STKTRKEAKKKNAYEEKKNAYOKANQAVLAKKASS-----YYIL----- 184
Qy 69 DGEWISGF--EGKIDAG-----YVINISKOTFIPVPEKIEKKEEENKPTFVSKKON 121
Db 185 -GWEEGGVGPVPHKKEENMLSHLYSSKOKENISKENDVLDE--KEEAEETEETEELBEKN 242
Qy 122 POWNHLSNREKEDLOREH-----SQKSDSTQDVATVLDKNISKSTTN 171
Db 243 EETESSEISEDEEEBEKEEKEQAKEQSENNNDQKKDMA-----QNLISKQNNN 295

```

## RESULT 24

```

Q9LH98 PRELIMINARY; PRT; 2081 AA.
ID Q9LH98
AC Q9LH98;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Genomic DNA, chromosome 3, BAC clone: T19N8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety pl.
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
DR EMBL: AP002057; BAB03174.1; -.
SQ SEQUENCE 2081 AA; 232851 MW; D3603EIF85EFFF29 CRC64;

```

Query Match 11.8%; Score 105.5; DB 10; Length 2081;



Best Local Similarity 25.4%; Pred. No. 50;  
Matches 48; Conservative 31; Mismatches 71; Indels 39; Gaps 8;

QY 2 IYVKDFARNTTKEFLINKDTGEVSELKPHRTVTIQTQNGKMSSTIVSEDFILPYKKE 61  
DB 1639 IYERNKGKEISKE-----GSEIDGK-----TPEINGSEBELSTERGSKD-----GK 1678  
QY 62 LKGVGFDPDMEISGEGEKKD-----AGVYINLSKDTFLKPVFKKIEEKEENK 111  
DB 1679 IEKGK--EKGKENTKESKDDKIEEGEKGKENTKESKDDKINEITHGDKKATMEGSKD 1736  
QY 112 --TFPYSK-KKDPQVNHSQLNESHKREDIQ-----REHSOKSDSTKDYATATVLDKNIS 164  
DB 1737 GGINSGKSKDSKSKSVINQVNDLSKDSKNDINEINNGKDSYKDNVTEIQGNDNL 1796  
QY 165 SKSTNNPN 173  
DB 1797 TNSTSEPN 1805

## RESULT 25

Q8IDA0 PRELIMINARY; PRT; 3127 AA.  
AC 08IDA0;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, last sequence update)  
DE 01-MAR-2003 (TREMBlrel. 23, last annotation update)  
DE Guanylyl cyclase (EC 4.6.1.2).  
GN GC-BETA.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
ON NCBI\_Taxid=36329;  
RN [1]  
RA SEQUENCE FROM N.A.  
RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,  
RA Bertrman M., Pain A., Hall N., Ackin R., Chillingworth C., Doggett J.,  
RA Omond D., Sanders M., Hayes R., Hall S., Hall M., Barrell B.,  
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL844509; CAD52725.1; -.  
KW Lyase.  
SQ SEQUENCE 3127 AA; 370597 MW; F0375C72B9007560 CRC64;

Query Match 11.7%; Score 105; DB 5; Length 3127;

Best Local Similarity 23.7%; Pred. No. 84;  
Matches 55; Conservative 36; Mismatches 73; Indels 68; Gaps 10;

QY 9 RNTTVK-EFLINKDTGEVSELK-----HRTVTIQTQNG-----KEMSTIVSE 50  
DB 1950 RNTTKKSDYLDLDNGESKFKFRNTSYVLESPLHLIGDIVNNIKRKKKKEIKTIVSD 2009  
QY 51 EDFTLPYKGELEKGYQFGWEISG--FEKKDAGVYNLSKDTFLKPVFKKIEEKEE 107  
DB 2010 DMFTSPVNIKEYNNEQERKEIVGNLSYDKTKKIPFIFTEGRIKK--KKIEKKEK 2067  
QY 108 ENK-----PTF-----DVSKKDNQVNHSQL----- 129  
DB 2068 EKEENNNFLYNDYSSYSPKTDGNNENFVITIERKDPQKKFHPNPFKFLHNTY 2127  
QY 130 ---NESHK---EDLQREHSOKSDSTKD-VTATVLDKNISSKSTNNENK 174  
DB 2128 PMNKKNKKKNKNKNVNRNRYNPTTSSKQDVSNPLSDLSFFSDNEYSQNE 2179

## RESULT 26

Q94C59 PRELIMINARY; PRT; 540 AA.  
AC 094C59;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)  
DE Putative phosphatidylinositol-transfer protein.  
GN T518.14.  
OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
ON NCBI\_Taxid=3702;

QY 11 TTVKEFLINKDTGEVSELKPHRTVTIQTQNGKMSSTIVSEDFILPYKGELEKGYQPDG 70  
DB 27 TTVKA-VVEETVVEDESKP-----EGVEKSAFKEESDFADLKESEKK----- 70  
QY 71 WEISGFEKKDAGVYNLSKDTFLKPVFKK--IEKKEENKPPFDYSKKDNQVNH 127  
DB 71 -LMSDKSLTEBAIVDN---TLTKTKKESSPMKKEEVKPAPEVKKKE--EAAEE 123  
QY 128 QNESHKREDLQREHSOKSDSTKDYATATVLDKNISS 165  
DB 124 KVEEKKSEAVTEAPKAEVAVTEETIIRKEEYTT 161

Query Match 11.6%; Score 104; DB 10; Length 540;

Best Local Similarity 25.3%; Pred. No. 15;  
Matches 40; Conservative 28; Mismatches 64; Indels 26; Gaps 6;

QY 11 TTVKEFLINKDTGEVSELKPHRTVTIQTQNGKMSSTIVSEDFILPYKGELEKGYQPDG 70  
DB 27 TTVKA-VVEETVVEDESKP-----EGVEKSAFKEESDFADLKESEKK----- 70  
QY 71 WEISGFEKKDAGVYNLSKDTFLKPVFKK--IEKKEENKPPFDYSKKDNQVNH 127  
DB 71 -LMSDKSLTEBAIVDN---TLTKTKKESSPMKKEEVKPAPEVKKKE--EAAEE 123  
QY 128 QNESHKREDLQREHSOKSDSTKDYATATVLDKNISS 165  
DB 124 KVEEKKSEAVTEAPKAEVAVTEETIIRKEEYTT 161

RESULT 27  
Q9SA84 PRELIMINARY; PRT; 540 AA.  
AC 09SA84;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)  
DE T518.14 protein (Hypothetical protein).  
GN T518.14 OR AT1G30650.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.  
ON NCBI\_Taxid=3702;

[1]  
RN [1]  
RA SEQUENCE FROM N.A.  
RA STRAIN=CV. Columbia.  
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,  
RA Li J., Kremenetskaia I., Luros J., Ngan I., Gonzalez A., Altafi H.,  
RA Araujo R., Chao Q., Conn L., Conway A.B., Dunn P., Hansen N.,  
RA Huizar L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,  
RA Davis R.W., Ecker J.R., Federpiele N.A., Theologis A.,  
RT "Arabidopsis thaliana chromosome 1 BAC T518 sequence."  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

[2]  
RN [2]  
RA SEQUENCE FROM N.A.  
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan Y.W., Lee J.M.,  
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,  
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,  
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,  
RA Ecker J.R., Theologis A.,

Query Match	11.5%	Score 103.5	DB 5	Length 325
Best Local Similarity	23.8%	Pred. No. 9.5		
Matches	43	Conservative	25	Mismatches 50, Indels 63, Gaps 9

  

QY	31	HRVTVTIONGKEMSTSI	SEEDFILPVYK	-GEL-	EKGQFDGWEISGE	EGK-----	79
		:	:		:	:	
DB	57	HRTTISIKN--RFSYK	KGIDDEEKLPRIS	KNGELI	VLNEL	EPDNFHIK--	EGKHLRKS
		:	:	:	:	:	:
QY	80	---KQAGV-----	INLSKDTFIKPV-		-FKK-----		100
		:	:	:	:		
DB	113	NHIKSGYVYANNEEL	IFLESCTLKEIT	LAQTKR	NSYKRN	TIINKLP	EEEEEEEEEE
		:	:	:	:	:	:
QY	101	--IEEKKSEENKPT	VDVSKKDN	PQVNSQL	NESSHKE-----	DLQREHS	SOKSDSYK
		:	:	:	:	:	:
DB	173	EEEEQGEVEVEK	APTISEEEBET	PAVSEB	EKKEEEEB	ETPAVSEB	EKKEEEEQEDBKX
		:	:	:	:	:	:
QY	152	P	152				
DB	233	D	233				

RESULT	30
093424	
ID	093424
AC	093424, P80801, PRELIMINARY; PRT, 385 AA.
DT	01-JAN-1999 (TrEMBLrel. 09, Created)
DT	01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE	Hypothetical glycine-rich 37.0 kDa protein E02A10.2 in chromosome V precursor.
DE	EN02A10.2
OS	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC	Rhabditidae; Peloderinae; Caenorhabditis.
OX	NCBI_TaxID=6239;

```

RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Thomas K.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP REVISIONS.
RC STRAIN=BRISTOL N2;
RA Jones S.J.M.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; 281053; CAB02877.1; -.
DR HSSP; P10968; 2CWC.
DR WormRep; E02A10.2; CE09116.
DR InterPro; IPR002952; Eggshell.
DR PRINTS; PR01228; EGGSHELL.
KW Hypothetical protein; Signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 385
FT TRANSMEM 165 185
FT DOMAIN 23 190
FT SEQUENCE 385 AA; 36963 MW; 32AB25AE9B5073FE CRC64;
SQ

Query Match 11.4%; Score 102.5; DB 5; Length 385;
Best Local Similarity 26.2%; Pred. No. 13;
Matches 48; Conservative 31; Mismatches 55; Indels 49; Gaps 9;

QY 1 KIVKDFPANTTKE-----TYKEFLINQD-----TGEVSELPKPRVTVTIO- 38
DB 215 RILNNTKTEVASEIKTLKEKAGDYIVVCKNEKAPFAETDPCSLQKENVHCTIIR 274
QY 39 -NKGEMSTIVSEDFILPYKGELEKGYQFDGMEISGFEKKDAGVYINLSKDTFLKPV 97
DB 275 INKEVAER--NEEDKKEEPPKKEEKKKEVEKKE--EDEKQD-----EP- 316
QY 98 FKIEEKKEEENPTDVSKKKQNPQVNSQLNESHKEDLQREHSSQKSDSTKDYATVY 157
DB 317 -KEEKEKEEKE--EVEKKE-----EKKQDEPKKEEKEEKEKEDVEKSEKV 368
QY 158 LDK 160
DB 369 EEK 371

RESULT 31
Q8IHW3 PRELIMINARY; PRT; 449 AA.
AC O8IHW3;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PFI_0413.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_TaxID=36329;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RC MEDLINE=22255705; Pubmed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlson J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyte S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angioli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Valdivia A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA MvFadden G.I., Cummings I.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Garucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.;
RT "Genome sequence of the human malaria parasite Plasmodium
falciparum."
RL Nature 419:498-511(2002).
DR EMBL; AE014842; AAN35996.1; -.

```

```

KW Hypothetical protein.
SQ SEQUENCE 449 AA; 52812 MW; 85B62272D6257C68 CRC64;

Query Match 11.4%; Score 102.5; DB 5; Length 449;
Best Local Similarity 24.9%; Pred. No. 16;
Matches 52; Conservative 30; Mismatches 58; Indels 69; Gaps 9;

QY 1 KIVKDFPANTTKE--FILNKDTGSEVSELPKPRVTVTIO NKGEMSTIVSEDFILPY 58
DB 31 KSVRDVADPLFVSESYSENK-----KELLTKLKEK--ILILY 70
QY 59 -----KGELEKGYQFDGMEISGFEKKDAGVYINLSKDTFLKPVFKIF----- 102
DB 71 NEQONEEKKKKKKNFQSD--ISEKKKSKADLQKSESNNNLKKKKRIESEEETNINS 127
QY 103 --EKKEE-----NKPTFDV-----SKKQNPQVNSQLNESHKEDLQREH 143
DB 128 DDEEEEEYORKKQKQKNSVSTLSLEKKKKRRDSESSNNNDYNEIYEDDEQEE 187
QY 144 SOKSDSTKDVATVLDKNNISSKSTTNP 172
DB 188 EEEES-----LSKKSKKKKNTTSP 207

RESULT 32
Q98QAL PRELIMINARY; PRT; 622 AA.
AC Q98QAL;
DT 01-OCT-2001 (T-EMBLrel. 18, Created)
DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE LIPOPROTEIN.
GN MYPV 4650.
OS Mycoplasma pulmonis.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
ON NCBI_TaxID=2107;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAB CTIP;
RC MEDLINE=21267165; Pubmed=11353084;
RA Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Gallison F.,
RA Moszer I., Dydwig K., Wroblewski H., Viart A., Kocha E.P.C.,
RA Blanchard A.;
RT "The complete genome sequence of the murine respiratory pathogen
Mycoplasma pulmonis."
RL Nucleic Acids Res. 29:2145-2153(2001).
DR EMBL; AL445564; CAC13638.1; -.
DR Mypulist; MYPV_4650; -.
KW Complete proteome.
SQ SEQUENCE 622 AA; 73762 MW; 42BD8930861960D CRC64;

Query Match 11.4%; Score 102; DB 16; Length 622;
Best Local Similarity 24.9%; Pred. No. 24;
Matches 48; Conservative 35; Mismatches 66; Indels 44; Gaps 11;

QY 9 RNTTYKEPIL-NKDTGSEVSELP-----KPRVY---TVTIOG--KEM--SPT--V 48
DB 26 KDSAKNFVFKSSDLISIEKQFKKNNLDKQKIIEGVITFINSETIEYIKETIILNL 85
QY 49 SEEDFILPYKGELEKGYQFDGMEISGFEKKDAGVYINLSKDTFLKPVFKIEEKEE 108
DB 86 SEDNIFISLNTENNNAKFLDLEF-----VSDEKFKIKFQGIINSQTBO 129
QY 109 NKPTDVSKKKD--NPQVNSQLNES--HKEDLQREHSSQKSDSTKDYATVLDK-- 161
DB 130 -KITDNISSKEDEKKNPNQNNNSNNSSDQKNDLQKNSDKLNDVQDEKANKNSNS 188
QY 162 NTSKSTTNPNK 174
DB 189 NDSKEKNDENTNK 201

RESULT 33

```

Q9UBD4  
ID Q9UBD4 PRELIMINARY; PRT; 1419 AA.  
AC Q9UBD4;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE S-adenosylmethionine decarboxylase-ornithine decarboxylase.  
OS Plasmodium falciparum.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=5833;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20179918; PubMed=10713131;  
RA Mueller S., De'dara A., Luessen K., Wrenger C., Das Gupta R.,  
RA Madhubala R., Walter R.D.;  
RT "In the Human Malaria Parasite Plasmodium falciparum, Polyamines Are  
RT Synthesized by a Bifunctional Ornithine Decarboxylase, S-  
RT Adenosylmethionine Decarboxylase.";  
RL J. Biol. Chem. 275:8097-8102 (2000).  
DR EMBL; AF094833; AAF00073.1; -.  
DR HSRP; P07805; 1F3T.  
DR InterPro; IPR000183; Decarboxylase2.  
DR InterPro; IPR001985; SAM decarbox.  
DR Pfam; PF02784; Orn\_Arg\_dec\_N; 1.  
DR Pfam; PF00278; Orn\_DAP\_Arg\_dec; 1.  
DR Pfam; PF01536; SAM\_decarbox; 1.  
DR PRINTS; PR01179; ODADCRBLASE.  
DR ProDom; PD002379; SAM\_decarbox; 1.  
SQ SEQUENCE 1419 AA; 166441 MW; 8842897C05056EE CRC64;

Query Match 11.4%; Score 102; DB 5; Length 1419;  
Best Local Similarity 22.0%; Pred. No. 58; Matches 77; Indels 70; Gaps 11;  
Matches 52; Conservative 37; Mismatches

QY 4 VQDFARNTTKEFIINKDTGEVSEL-----KPRVTYTIQNGKMSSTIYSEEDPT 54  
DB 1019 VEDMSNNGFNFIYIINLGSGYPELELYDAKGDHDKHYTSLDLQKDKIOKFLNEFTL 1078  
QY 55 LEVY-----KGELEKGYOFDMEISGFEGKKD-----AGYINISKDTFIKVPKK 100  
DB 1079 KTKYGYGFSEKISLAINMSIDHY-----FSHMDNLRVICEPORYWVAASSTLAVKIKGR 1134  
QY 101 -----IEEKKKEENKPTF--DYSKKDNPQVNH--SQNESHRKEDLOREHSQK- 146  
DB 1135 RPTFOGIMLKKELKDHYPDLNFAQGENKQKQDETKIHNNNDNNNNNNNNNNNNNNNOGK 1194  
QY 147 -----SDST-----KDVATVLDK--NNIS-SKSTNNPN 173  
DB 1195 GQGNIMNDLITSTDSISKNDHSSQYIQNVSCITIRKEDNLIKINHTINNPN 1250

RESULT 34  
Q8IFNO PRELIMINARY; PRT; 1612 AA.  
AC Q8IFNO;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
CN PF0115C.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hamlin N., Pain A., Berriman B., Hall N., Bowman S., Churcher C.,  
RA Harris B., Harris D., Lawson D., Quail M., Bartell B.,  
RA Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AL034477; CAD49270.1; -.  
DR Hypothetical protein.  
SQ SEQUENCE 1612 AA; 190946 MW; F00A408EB4521D07 CRC64;

Query Match 11.4%; Score 102; DB 5; Length 1612;

Best Local Similarity 27.2%; Pred. No. 67;  
Matches 47; Conservative 29; Mismatches 73; Indels 24; Gaps 9;

QY 9 RNTYKFEITLNGDGEVSELKPRVTYTIQNG-KEMSTIYSE-----EDFLPYKKEL 62  
DB 1154 RNSMNQNIINNKTDIKFCKENDKFKYKIKIMFTPYKELNLYDGFNSNHNML 1213  
QY 63 EKGVOFDGMEISGFEGKDGAVINLSKDTF-IKVPFKIEKKEEKPKFPDVSCKKON 121  
DB 1214 AKESWYTW-----KKQIN-VAKINQDYDISFKFCIDENIKQN---YDINQTTNN 1262

QY 122 PQVNSQNESHRKEDLOREHSQKSDTKDVTATVLDKNNISSKSTNNPNPK 174  
DB 1263 NTYNNQTSN--HMENLHFTDE-SSKSNNNQDKTYVT---NEWSYLPNIPKQKX 1309

RESULT 35  
Q9PPL5 PRELIMINARY; PRT; 312 AA.  
ID Q9PPL5  
AC Q9PPL5;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Putative membrane protein.  
GN CJO692C.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;  
OC Campylobacteriaceae; Campylobacter.  
OX NCBI\_TaxID=197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCITC 11168;  
RX MEDLINE=20150912; PubMed=10688204;  
RA Parthill J., Wren B.W., Mungall K., Kelley J.M., Churcher C.,  
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,  
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,  
RA Quail M.A., Raftery M.A., Rutherford K.M., van Vliet A.H.M.,  
RA Whitehead S., Bartell B.G.;  
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni  
RT reveals hypervariable sequences.";  
RL Nature 403:665-668 (2000).  
DR EMBL; AL139076; CAB72966.1; -.  
KW Complete proteome.  
SQ SEQUENCE 312 AA; 37221 MW; 0004FA7936A7418B CRC64;

Query Match 11.3%; Score 101.5; DB 16; Length 312;  
Best Local Similarity 25.0%; Pred. No. 13;  
Matches 46; Conservative 34; Mismatches 73; Indels 31; Gaps 9;

QY 15 EFLNKDTGEVSELKPRVTYTIQNGKMSSTI--VSEDFILPYK-----GELE 63  
DB 5 DFIKERQNIROKMLKFSR---AINQKPLDDLRDEISSDLIRRRFKKTPNKLEIED 61  
QY 64 KGYQFDGMEISGFEGKQAGVINL---SKDTFIKVPFKIEEKKKEENKPT---FDVSK 117  
DB 62 EYESKHTYKKSNTIYKED--LVNKLKEKOSLAKKIFSCKMERKEENKKTKNFLFSR 118  
QY 118 KKDNP-----QVNSQNESHRKEDLOREHSQKSDTKDV--TATVLDKNNISSK--STT 169  
DB 119 KKANEIKNIQTKQTSNQAATTQTKQEKKEELNINSIEKIKTEKTKIQKPLIEKKDLVK 178

QY 170 NNPN 173  
DB 179 NQPN 182

RESULT 36  
Q9V7J0 PRELIMINARY; PRT; 382 AA.  
ID Q9V7J0  
AC Q9V7J0; Q9G081;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

CG8421 protein (Aspartyl beta-hydroxylase variant 2).  
 ASPH OR CG8421 OR CG18658.  
 GN Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RC STRAIN=Berkley;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Ballaw R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Houtin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matvei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merilov G., Mishina N.V., Mobarry C., Morris J., Mostreli A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson W., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN  
 RN SEQUENCE FROM N.A.  
 RA Ceiniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorett V., Doup L.E., Doyle C., Dresnek D., Fafan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hoshin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Matvei B., Mostreli A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacle J., Pargass V., Park S., Patel S., Pfeiffer B.,  
 RA Phouenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN  
 RN SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradsky P., Huang Y., Kaminck J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Kaminck J.S., Carlson J.W., Ceiniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,

RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN  
 RN SEQUENCE FROM N.A.  
 RA Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN  
 RN SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=20564328; PubMed=10956655;  
 RA Dinkov J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
 RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
 RA Friedman P.A.;  
 RT "Aspartyl beta-Hydroxylase (Asph) and an Evolutionarily Conserved  
 RT Isoform of Asph Missing the Catalytic Domain Share Exons with  
 RT Junction.";  
 RL J. Biol. Chem. 275:39543-39554(2000).  
 DR EMBL; AF003808; AA058063.2; -;  
 DR EMBL; AF289494; AAG40807.1; -;  
 DR FlyBase; FBgn0034075; Asph.  
 SQ SEQUENCE 382 AA; 43287 MW; 60B5C03AEBFC6E8B CRC64;  
 Query Match 11.3%; Score 101.5; DB 5; Length 382;  
 Best local Similarity 24.5%; Pred. No. 16;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 QY 51 EEFILPVYGELEKGYQFDGW-----EISGEGKKDAGYVI-----NLSKDTFK 95  
 DB 78 EDLDITPLSRKSK-VFGGWDEHRDEHDGVDGPSEALDDHDHDDHDEDEDE 135  
 QY 96 PVFKIEEKKEENKPT-----FVSKKKDNPQVNHQSINSHKREDIQREHSQKSDS 149  
 DB 136 PLTELEEELEEEBETEDDEPADEVEDEDEENNA--GENITAEDEEEEBEDND 193  
 QY 150 TKDVTATVLDKNISKST 168  
 DB 194 EGTVEATVATTEATTEAT 212  
 RESULT 37  
 Q17595 PRELIMINARY; PRT; 535 AA.  
 AC Q17595;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical 60.3 kDa protein.  
 GN C02H7.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 NCBI\_TaxID=6239;  
 RN  
 RN SEQUENCE FROM N.A.  
 RA STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RT None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN  
 RN SEQUENCE FROM N.A.  
 RA STRAIN=Bristol N2;  
 RA Leinbach D., Minx M.;  
 RT "The sequence of C. elegans cosmid C02H7.";  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 RN  
 RN SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission."  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U49945; AAC7924.1;  
 DR WormPep; C02H7.1; CE06757.  
 KW Hypothetical protein.  
 SQ SEQUENCE 535 AA; 60303 MW; 03B2DBE8BA3DFFB6 CRC64;  
 Query Match 11.3%; Score 101.5; DB 5; Length 535;  
 Best Local Similarity 21.0%; Pred. No. 22;  
 Matches 39; Conservative 31; Mismatches 71; Indels 45; Gaps 6;  
 QY 4 VQDFARNTTVEKFLNK-----DTGEVSLKPHRVTVTQNGKMSSTIVSEEDTILPY 58  
 DB 55 IKSAGNDKNTKTAFLDKLIKIDDSLKNVKAkti-----SGKDABET----- 98  
 QY 59 KGELEKGYQFDGMEISGEGKKDAGVYNLSKDTFIKPVFKIEEKEEENKPTFDVSK 118  
 DB 99 ----NKMQLMGTNATSFNSRNGTG-----EKKKKKKKKEDKKGDEEEST---TKK 144  
 QY 119 KDNPNVNSQLNESHKEDLQREHSQKSDSTK-----DYATATVLDKNNISKST 168  
 DB 145 RSSKKEETHEKSEKSEKSAEKEKSSSSSKERHKSSDRSSEKSEKSEKSEKST 204  
 QY 169 TNNPK 174  
 DB 205 TDEKPK 210  
 RESULT 38  
 Q9V719 PRELIMINARY; PRT; 556 AA.  
 AC 09V719;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE CG6421 protein.  
 GN ASPH OR CG6421 OR CG18658.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCB1\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkelley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Beus P.V., Bertan B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Broctier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milhina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter B., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.C., Wu D., Yang S., Yao O.A.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Dou P.L.E., Doyle C., Dresnek D., Fartan D.,  
 RA Ferreira S., Frise E., Galie R.F., Gary N.S., George R.A.,  
 RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jaitani M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman Y., Bertan B., Carlson J.W., Celniker S.E.,  
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Seale S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003808; AAF58064.2;  
 DR FlyBase; FBgn0034075; Asph.  
 SQ SEQUENCE 556 AA; 63144 MW; B420980CBDC6357A CRC64;  
 Query Match 11.3%; Score 101.5; DB 5; Length 556;  
 Best Local Similarity 24.5%; Pred. No. 23;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 QY 51 EDLILPVKGLGEGYQFDG-----EISGFEKKDAGVYNLSKDTFIK 95  
 DB 78 EDLILPVKGLGEGYQFDG-----EISGFEKKDAGVYNLSKDTFIK 95  
 QY 96 PVFKIEEKEEENKPT-----PDVSKKDNPNVNSQLNESHKEDLQREHSQKSDS 149  
 DB 136 PVFKIEEKEEENKPT-----PDVSKKDNPNVNSQLNESHKEDLQREHSQKSDS 149  
 QY 150 TKQYATATVLDKNNISKST 168  
 DB 194 EGTVEATVATTEATTEAT 212  
 RESULT 39  
 Q9G082 PRELIMINARY; PRT; 765 AA.

AC 09G082;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Aspartyl beta-hydroxylase variant 1 (CG8421-PA).  
 NC ASPH OR CG8421 OR CG18658.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_Taxid=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20564328; PubMed=10956665;  
 RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,  
 RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,  
 RA Friedman P.A.;  
 RT "Aspartyl beta -Hydroxylase (Asph) and an Evolutionarily Conserved  
 RT Isoform of Asph Missing the Catalytic Domain Share Exons with  
 RT Junction";  
 RL J. Biol. Chem. 275:39543-39554(2000).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazel R.G., Champe M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtsky K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostel C., Gabelian A.E., Garg N.S., Gelbart J.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kesterson J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murray D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Switzkae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.W., Weisenbach J.,  
 RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amaratunga P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorett V., Dou P.L.B., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galie R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Paclab J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Switzkae R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Bertram B., Carlson J.W., Celinker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Seattle S.M.J., Smith E., Shu S., Smurniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [5]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 [6]  
 RP SEQUENCE FROM N.A.  
 RA Flybase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF289493; AAG40806.1;  
 DR EMBL; AE003808; AAM70947.1;  
 DR Flybase; FBgn0034075; Asph.  
 DR InterPro; IPR001440; TPR.  
 DR InterPro; IPR006025; Zn\_MTPetase.  
 DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
 SO SEQUENCE 785 AA; 89843 MW; 30A8DFCD636F7F1 CRC64;  
 Query Match 11.3%; Score 101.5; DB 5; Length 785;  
 Best Local Similarity 24.5%; Pred. No. 34;  
 Matches 34; Conservative 29; Mismatches 51; Indels 25; Gaps 5;  
 DB 51 EPIILPVYKGELEKGVQFPGW-----EISGEGKKDAGVY-----NLKSDTFIK 95  
 78 EILDPPLSSRSRK--VFDGWVDEHDEHDGVDVPSGEALDDHDDHDDHDEDEDE 135  
 QY 96 PVFKKIEEKEEENKPT-----PVSKKKQPNQVHNSQJNSHREKDLQREHSOKSDS 149  
 DB 136 PLTEHELELEEEBTEDEDEPADEVEYDEDEENNA--GENITAEDEEEEBEDND 193  
 QY 150 TRDYATATVLDKNNISKST 168  
 DB 194 EGTVEATVTEATTEAT 212  
 RESULT 40  
 ID Q81556 PRELIMINARY; PRT; 1377 AA.  
 AC Q81556;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Eukaryotic translation initiation factor 3 subunit 10, putative.  
 GN PFL0625C.  
 OS Plasmodium falciparum (Isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 NCBI\_Taxid=363297.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=3D7;  
 RX MEDLINE=22255705; PubMed=12368864;  
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Chan M.-S., Nene V., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Eissen J.A., Bhat V., Shallow S.J., Sub B., Peterson J., Angiolini S.,  
 RA Perica M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,





**This Page Blank (uspto)**